

**Estudi de la variació del clima
marítim extremal a conseqüència
del canvi climàtic**

Treball realitzat per:
Raquel Feixa Compte

Dirigit per:
Maribel Ortego Martínez
Jesús Corral López

Màster en:
Enginyeria de Camins, Canals i Ports

Barcelona, juny de 2016

Departament d'Enginyeria Civil i Ambiental

TREBALL FINAL DE MÀSTER

Índice general

Lista de figuras	5
1. Objetivos	12
2. Antecedentes	14
3. Metodología	17
4. Zona de estudio	20
4.1. Procedencia y obtención del conjunto de datos	21
4.2. Parámetros disponibles	22
5. Caracterización del oleaje	23
5.1. Distribución conjunta de periodo de pico y altura significativa . .	24
5.2. Distribución conjunta de dirección y altura significativa	26
5.3. Distribución de Weibull por direcciones	28
5.4. Distribución Generalizada de Pareto	30
6. Datos composicionales	32

6.1. Definición de categorías	32
6.2. Análisis composicional de datos	35
6.3. Dendrograma de la partición	38
7. Estimación de los parámetros	40
7.1. Estimador de máxima verosimilitud de la distribución multinomial	40
7.2. Estadística bayesiana	41
7.2.1. Modelo en el espacio símplex	42
7.2.2. Modelo en el espacio real	42
7.3. Método de Hamilton Monte Carlo	43
8. Regionalización de la altura de ola	46
8.1. Variogramas.	46
8.2. Análisis cluster	48
9. Resultados	49
9.1. Regionalización de los datos originales	49
9.2. Regionalización de las muestras de las distribuciones posteriori . .	51
9.3. Muestras, en coordenadas, de las distribuciones posteriori	53
10.Conclusiones	61
11.Trabajos futuros	63
12.Agradecimientos	64

ÍNDICE GENERAL

A. Histogramas de las muestras posteriori	68
---	----

Índice de figuras

3.1. Código de signos de la SBP utilizada en el análisis de datos. . . .	18
4.1. Nodos SIMAR. (Fuente: Propia basada en Google Maps).	21
5.1. Histogramas de Tp y Hs.	24
5.2. Tabla Periodo de Pico - Altura Significante.	24
5.3. Distribución conjunta de Hs-Tp.	25
5.4. Rosa de Oleaje.	26
5.5. Histogramas de Altura Significante y Dirección.	27
5.6. Tabla Altura Significante - Dirección.	27
5.7. Serie de datos de altura de ola.	28
5.8. Distribuciones de Weibull.	29
5.9. Parámetros de las distribuciones de Weibull por direcciones. . . .	29
5.10. Función de probabilidad acumulada de la distribución generalizada de Pareto.	30
5.11. Función de densidad de la distribución generalizada de Pareto. . .	31

6.1. Serie de datos de altura de ola. Separación de 72h entre temporales independientes.	33
6.2. Datos clasificados por categoría y año.	34
6.3. Dendrograma donde se representa la media y varianza de las coordenadas. Los valores medios de los balances están representados en el dendrograma por los puntos de enlace de las líneas horizontales con las verticales. La varianza queda representada por los diagramas de caja.	39
8.1. Variogramas de los log-ratios de cada par de variables (i, j) de los datos composicionales.	47
9.1. Dendrograma de los 4 clústers obtenidos de los datos composicionales iniciales.	50
9.2. Regionalización en 4 clases a partir de los datos originales. (Fuente: elaboración propia basada en Google Maps)	50
9.3. Dendrograma de los 5 clústers obtenidos a partir de las muestras del posteriori. Las líneas vericales representan en nivel de disimilitud entre grupos.	52
9.4. Regionalización en 5 clases a partir de la muestra simulada. (Fuente: elaboración propia basada en Google Maps)	52
9.5. Histogramas de la coordenada 1 del nodo SIMAR 2132124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	54
9.6. Histogramas de la coordenada 2 del nodo SIMAR 2132124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	54
9.7. Histogramas de la coordenada 3 del nodo SIMAR 2132124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	55
9.8. Histogramas de la coordenada 4 del nodo SIMAR 2132124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	55

ÍNDICE DE FIGURAS

9.9. Histogramas de la coordenada 1 del nodo SIMAR 2100124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	56
9.10. Histogramas de la coordenada 2 del nodo SIMAR 2100124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	56
9.11. Histogramas de la coordenada 3 del nodo SIMAR 2100124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	56
9.12. Histogramas de la coordenada 1 del nodo SIMAR 2122130. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	57
9.13. Histogramas de la coordenada 2 del nodo SIMAR 2122130. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	57
9.14. Histogramas de la coordenada 3 del nodo SIMAR 2122130. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	58
9.15. Histogramas de la coordenada 1 del nodo SIMAR 2132148. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	58
9.16. Histogramas de la coordenada 2 del nodo SIMAR 2132148. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	59
9.17. Histogramas de la coordenada 3 del nodo SIMAR 2132148. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	59
9.18. Histogramas de la coordenada 1 del nodo SIMAR 2090124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	60
9.19. Histogramas de la coordenada 2 del nodo SIMAR 2090124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	60
A.1. Histogramas de la coordenada 4 del nodo SIMAR 2100124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	68
A.2. Histogramas de la coordenada 4 del nodo SIMAR 2122130. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	69
A.3. Histogramas de la coordenada 4 del nodo SIMAR 2132148. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	69

A.4. Histogramas de la coordenada 3 del nodo SIMAR 2090124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	69
A.5. Histogramas de la coordenada 4 del nodo SIMAR 2090124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.	70

(Nota: Las figuras en las que no se especifica la fuente son de elaboración propia.)

Resumen

Palabras clave: datos composicionales, cambio climático, clima extremal, estadística bayesiana, simulación de Monte Carlo, clúster.

El objetivo principal de este estudio es detectar si la intensidad de los temporales extremos en la costa catalana ha aumentado debido al cambio climático durante los últimos 60 años. Para ello se dispone de datos de diferentes puntos SIMAR durante el periodo de tiempo 1958-2015, proporcionados por Puertos del Estado. Para afrontar el problema primero hemos definido 5 categorías de temporales y los hemos clasificado con el objetivo de hacer un tratamiento composicional de los datos de altura de ola. Hemos desarrollado un método de estimación de proporciones basado, por una parte, en la geometría de Aitchison en el simplex y en la correspondiente transformación *ilr* a coordenadas y, por otra, en técnicas de inferencia bayesianas mediante el uso del método de Hamilton Montecarlo.

Después de clasificar los temporales en categorías se dispone de datos composicionales y para trabajar en el espacio real hemos transformado los datos del simplex a coordenadas *ilr*. Se dispone de pocos datos y el número de temporales extremos es bajo, en algunos casos nulo. Por estos motivos es más adecuado utilizar métodos de estadística bayesiana por las ventajas que tiene frente los métodos clásicos. Un inconveniente de la estadística bayesiana es que la expresión de la función *posteriori* es complicada. Las técnicas bayesianas combinadas con el método de Hamilton Montecarlo nos han permitido obtener una muestra de esta distribución *posteriori*, mediante el uso de RStan. El proyecto Stan es una herramienta desarrollada recientemente que implementa inferencia estadística bayesiana y aporta calidad a las estimaciones por el algoritmo que utiliza.

Con la información *a posteriori* obtenida de los 15 nodos SIMAR hemos definido clústers utilizando un método clúster jerárquico para agrupar los nodos con mayor correlación y teniendo en cuenta la situación espacial de cada nodo. El análisis clúster se ha realizado utilizando la distancia Mahalanobis y los variogramas de los log-ratios de cada par de variables. Después de esta regionalización de los nodos hemos estudiado la evolución del clima extremal de un nodo de cada grupo.

Abstract

Key words: compositional data, climate change, extreme weather events, Bayesian data analysis, MCMC, cluster.

The main objective of this thesis is to analyse whether the intensity of extreme adverse weather conditions in the Catalan coast have increased as a consequence of the climate change over the last 60 years. The data used for this study come from different SIMAR points located in the Mediterranean coast for the period 1958-2015 provided by Puertos del Estado. The first step for this study has been the definition of 5 categories of storms, according to which we have classified the storms of each year. In order to estimate the probabilities of each category we have developed a method combining, on one hand, Aitchison geometry and the corresponding ilr transformation to coordinates, and on the other hand, Bayesian techniques using the Hamilton Montecarlo method.

The classification of storms between different categories provides us with compositional data, while the transformation of simplex data in ilr coordinates allows us to work in the real space. The number of extreme weather conditions is low, and in some cases zero and for that reason it is advisable to use Bayesian statistical methods for its advantages in front of other traditional methods. A common drawback of using Bayesian statistics is that the expression of the posterior function is complex. Nevertheless, the use of MC methods and RStan have allowed us to solve this issue and provide a sample of this distribution on the basis of which we have studied the evolution of extreme storms.

The SIMAR nodes have been grouped in clusters according to the information provided by the samples. We have used a hierarchic cluster method that allows to group the nodes with more correlation between them and taking into account the position of each node. The Mahalanobis distance and log-ratios variograms for each pair of variables have been used to define the clusters. Following the regionalization of the nodes, we have studied the evolution of one node of each group.

Capítulo 1

Objetivos

El objetivo principal consiste en determinar si se produce algún cambio en la intensidad de los temporales entre 1958 y 2015 en la costa catalana. Así pues, se desarrollará un método de estimación de proporciones, combinando técnicas bayesianas y métodos de Montecarlo, para estimar las probabilidades de los sucesos extremos de dos periodos de tiempo (1958-1988 y 1989-2015) y poder compararlos.

Con los datos proporcionados por Puertos del Estado en 15 nodos SIMAR de la costa catalana, primero se realizará un análisis estadístico previo basado en la obtención de histogramas, tablas de contingencia y rosas de oleaje. A continuación, se realizará la caracterización extremal de oleaje ajustando distribuciones de Weibull y la distribución Generalizada de Pareto.

Como el número de temporales por año es bajo se transformará la variable *altura de ola*, H_s , en una variable categórica, X . Se definirán cinco categorías de temporales y se pasarán los datos iniciales a datos composicionales. Esto nos permitirá hacer un tratamiento composicional de alturas de ola, basado en la geometría de Aitchison en el simplex y en la correspondiente transformación *ilr* a coordenadas. Teniendo en cuenta las características del problema, se dispone de pocos datos y el número de temporales de alguna categoría es nulo, el uso de los métodos clásicos de inferencia no es lo más adecuado. Por tanto, se realizará un enfoque alternativo al problema de inferencia sobre proporciones. La metodología que se utilizará se basa en el uso de técnicas bayesianas combinadas con los métodos de Montecarlo de simulación.

CAPÍTULO 1. OBJETIVOS

Seguidamente, con la información de las muestras simuladas de la distribución *a posteriori* y teniendo en cuenta la dependencia espacial entre los nodos, se clasificarán los 15 puntos SIMAR en clústers. Finalmente, comentaremos los resultados obtenidos acerca del comportamiento de un nodo de cada clúster.

Capítulo 2

Antecedentes

Para el diseño de obras marítimas es necesario conocer las acciones sobre la estructura, caracterizadas en función de la altura de ola, periodo y dirección. Estos datos de oleaje pueden ser instrumentales o procedentes de simulaciones numéricas. En este trabajo se utilizarán datos simulados proporcionados por Puertos del Estado. La variable que se utiliza comúnmente en el diseño de obras marítimas es la altura de ola significativa, definida como la media aritmética del tercio de olas más altas del conjunto de olas del registro. La probabilidad de que la altura de ola significativa del oleaje supere la altura de ola significativa de diseño, correspondiente a un periodo de retorno T , durante la vida útil de la obra, es lo que las ROM (ROM(2001)) definen como riesgo. Su máximo valor admisible viene fijado en función de las repercusiones económicas y la posibilidad de pérdidas humanas en caso de fallo de la estructura. Debido a la gran importancia de este parámetro en el diseño de obras marítimas el objetivo de este estudio es averiguar si ha habido un cambio en la distribución estadística de la altura de ola de los temporales, a consecuencia del cambio climático, durante los últimos 60 años. Se han realizado muchos estudios para cuantificar los efectos del cambio climático como el calentamiento global, los efectos en el clima marítimo, etc. A continuación, se presentan algunos estudios realizados durante los últimos años.

El estudio del clima es un campo de investigación en el que intervienen muchos parámetros como la temperatura y las precipitaciones. Las variaciones climáticas se han producido a lo largo de la historia de la Tierra y son debidas a causas naturales pero también a causas humanas. Por tanto, la investigación es fundamental para tomar medidas que minimicen estos impactos negativos de nuestro

modo de producción y consumo energético en el clima. En IPCC (2007) se ponía de manifiesto la evidencia del aumento de la temperatura en el siglo XX. Según este informe, en promedio, la temperatura ha aumentado aproximadamente 0,6°C en el siglo XX y el nivel del mar ha crecido de 10 a 12 centímetros.

En IPCC (2007) se concluye que la alteración de la frecuencia e intensidad de los fenómenos meteorológicos extremos, sumada al aumento del nivel del mar, tendrán previsiblemente efectos extremadamente adversos sobre los sistemas naturales y humanos.

En Casas-Prat & Sierra (2013) se pone de manifiesto la dificultad de saber como reaccionarán al cambio climático los patrones atmosféricos debido a las múltiples variables de las que dependen y las interacciones entre ellas y se concluye que la altura de ola significativa aumentará en un 10 % para las condiciones medias y un 20 % para el clima extremo.

Según Sánchez Arcilla & González-Marco (2008) en las últimas décadas el área Noroeste del Mediterráneo ha experimentado fuertes tormentas caracterizadas por grandes olas y vientos comparados con otros periodos. En Sánchez Arcilla & González-Marco (2008) se explica que el clima mediterráneo puede ser descrito como torrencial. Por este concepto se entiende un predominio de las tormentas de oleaje discontinuas comparativamente mucho más enérgicas que los periodos de calma adyacentes que se producen la mayor parte de las veces.

Según el informe Climate change adaptation, coastal and marine issues (2013), la frecuencia e intensidad de los eventos de clima extremo cambiarán como consecuencia del cambio climático. Las zonas costeras son particularmente vulnerables al incremento del nivel del mar y a la intensidad de los eventos extremos.

Según IPCC (2012) hay evidencias en las observaciones de un cambio en los sucesos extremos desde 1950 y se aborda el tema mediante la evaluación de la literatura científica sobre temas que van desde la relación entre el cambio climático y las condiciones meteorológicas extremas y los fenómenos climáticos extremos a las implicaciones de estos acontecimientos para la sociedad y el desarrollo sostenible. Se estudia la interacción de los factores climáticos, ambientales y humanos que pueden conducir a impactos y desastres, y se dan recomendaciones para reducir los riesgos del cambio climático.

En Harley *et al.* (2006) se estudian los impactos del cambio climático en los ecosistemas marinos y en los sistemas sociales y económicos que dependen de

ellos.

En Gisbert Martín de Hijas (2007) se concluye que se produce una variación en la distribución de la altura de los temporales alrededor del año 1995. En este trabajo se realizará un estudio similar con datos de un periodo de tiempo más extenso, de 1958 a 2015, y utilizando RStan para simular las muestras *a posteriori*. Stan utiliza el algoritmo de Hamilton Montecarlo que presenta ventajas respecto al de Gibbs utilizado en Gisbert Martín de Hijas (2007).

Capítulo 3

Metodología

El objetivo de este trabajo es estudiar si ha variado la probabilidad de los sucesos extremos como consecuencia del cambio climático. Como se disponen de pocos datos para ajustar una función de probabilidad de la variable aleatoria *altura de ola* se definirán categorías de temporales, en función de la altura de ola, y se transformará esta variable en una variable categórica. La variable transformada, X , cuenta el número de temporales de cada categoría registrados cada año y sigue una distribución multinomial con parámetro \mathbf{p} . Este vector \mathbf{p} pertenece al simplex, $S^D \subset \mathbb{R}^D$. Dado que existe una restricción de suma constante este subespacio tiene dimensión $D-1$.

S^D tiene estructura propia de espacio vectorial, con la distancia de Aitchison (Pawlowsky-Glahn *et al.* (2015)). Si se desea utilizar la estructura euclidiana usual, basta con transformar \mathbf{p} , expresando el vector en una nueva base de \mathbb{R}^{D-1} . Para obtener este vector es común utilizar la transformación $\text{ilr}(\mathbf{p})$ que permite definir una base ortonormal utilizando la llamada SBP (figura 3.1), en la cual se escogen los grupos de partes a comparar (Pawlowsky-Glahn *et al.* (2015)). Las coordenadas del vector transformado, θ , se pueden entender como balances entre grupos de partes de la partición. Nos interesa estudiar la evolución del balance entre los temporales poco intensos y moderados frente a los extremos, por ese motivo en la primera partición separamos las dos categorías más intensas del resto. Las expresiones de las coordenadas se especifican en el capítulo 6.

Balance/ Temporales	Débiles	Mod. débiles	Moderados	Mod. intensos	Intensos
1	1	1	1	-1	-1
2	0	0	0	1	-1
3	1	1	-1	0	0
4	1	-1	0	0	0

Figura 3.1: Código de signos de la SBP utilizada en el análisis de datos.

Como se indica en la figura 3.1, la partición SBP es:

$$[x_1|||x_2|||x_3|x_4|x_5]$$

Para estimar el vector de probabilidades se utilizarán técnicas bayesianas (Gelman *et al.* (2014)) ya que el número de temporales intensos en algunos casos es nulo y si se utilizaran métodos clásicos, como el estimador de máxima verosimilitud, la probabilidad de los eventos extremos daría cero, siendo esto falso. La estimación bayesiana se basa en suponer que los parámetros desconocidos no tienen un valor fijo sino que son variables aleatorias que siguen una densidad de probabilidad que modela la incertidumbre acerca de su valor. La distribución de probabilidad de los parámetros desconocidos, θ , dados los datos, es igual, utilizando el teorema de Bayes, a la probabilidad de los datos condicionada a unos valores de los parámetros fijos multiplicada por la distribución de los parámetros *a priori*:

$$p(\theta|x) = p(\theta)p(x|\theta). \quad (3.1)$$

La distribución *a priori* se basa en la información que se posee y expresa el grado de creencia que se tiene antes de estudiar los datos. Como en este caso se dispone de poca información *a priori*, la distribución que se utilizará es una normal multivariante objetiva. Para que la función *a priori* sea objetiva las varianzas de la matriz de covarianzas tienen que ser grandes. De esta forma la función *a posteriori* se parecerá a la verosimilitud de los datos y la mayor parte de la información que se aportará al modelo será la de los datos. La función de verosimilitud introduce la información muestral que contienen los datos, es la probabilidad que los datos sean generados por el modelo, dados unos valores de los parámetros. Para la función de verosimilitud de los parámetros en este caso se ha considerado una multinomial. Considerando que cada fila de la matriz de datos, los temporales de cada año, son independientes, la verosimilitud es el producto de la distribución multinomial aplicada a los datos de cada año.

Para trabajar en coordenadas se han escrito las probabilidades en función de las coordenadas utilizando la transformación *ilr* inversa. De esta forma se obtendrá la función de probabilidad multinomial, en coordenadas, que resulta después del cambio de variable del *símplex* al espacio real (Pawlowsky-Glahn *et al.* (2015)). La distribución conjunta *a posteriori* resultante tiene una expresión complicada, por tanto será necesario utilizar los métodos de Montecarlo (Robert & Casella (2004)) para simular una muestra de la distribución *a posteriori*, a partir de la cual se podrá realizar cualquier inferencia sin necesidad de integrar la función *posteriori*. El método de muestreo que se utilizará en este caso es el de Hamilton Monte Carlo (Gelman *et al.* (2014)), explicado en el capítulo 7, mediante el uso de RStan (ver Stan Development Team (2015)). La muestra simulada permitirá obtener información de las proporciones entre categorías de temporales. La muestra se obtiene en coordenadas que se pueden transformar al espacio *símplex* con la transformación *ilr* inversa.

Con esta información *a posteriori* los 15 nodos se han agrupado en *clústers* considerando los valores de las probabilidades de cada categoría y teniendo en cuenta la situación espacial de cada nodo. La dependencia espacial de los datos composicionales se ha caracterizado mediante *variogramas*, como se hace en Ortego *et al.* (2011) para los datos pluviométricos. Para definir los *clúster* se utiliza el método de Ward (Everitt (1993)). Inicialmente, cada objeto forma un grupo independiente y a continuación los objetos o grupos próximos se juntan sucesivamente. Para evaluar la correlación entre los nodos se define la distancia Mahalanobis utilizando los *variogramas* de los *log-ratios* de cada par de variables, que caracterizan la dependencia espacial de los datos. Este análisis *clúster* se ha realizado para los datos composicionales iniciales y para las muestras simuladas, en el segundo caso se ha obtenido un grupo más que en el primero. Una vez obtenidos los *clústers* se ha observado la evolución del clima extremal de un nodo de cada grupo, es decir de 5 nodos en total.

Capítulo 4

Zona de estudio

El objetivo de este estudio es comprobar si hay evidencias de cambio climático a partir del análisis estadístico de un conjunto de nodos que proporcionan información de la altura de ola de los temporales. Estos datos utilizados han sido facilitados por Puertos del Estado y corresponden a 15 nodos SIMAR durante el periodo 1958-2015. Estos nodos están situados en el litoral catalán y están señalados en verde en la imagen 4.1. Para cubrir un área más grande no solo se dispone de información de nodos a lo largo de la costa, sino también de otros situados en aguas profundas.

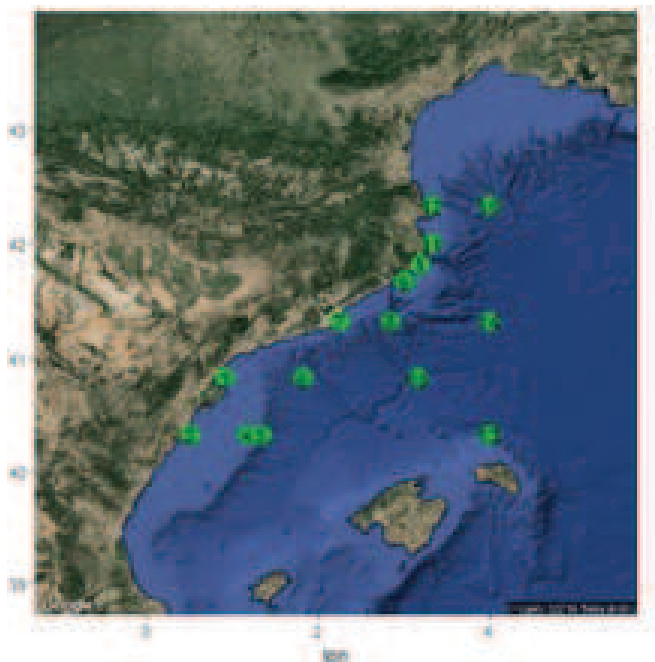


Figura 4.1: Nodos SIMAR. (Fuente: Propia basada en Google Maps).

4.1. Procedencia y obtención del conjunto de datos

El conjunto de datos SIMAR (ver Puertos del Estado (a)) está formado por series temporales de parámetros de viento y oleaje obtenidos mediante simulación numérica, no son datos instrumentales. Para poder ofrecer series temporales más extensas en el tiempo y actualizadas diariamente las series SIMAR combinan los grupos de datos simulados de oleaje SIMAR-44 y WANA. De este modo, el conjunto SIMAR ofrece datos desde el año 1958 hasta la actualidad.

El conjunto SIMAR-44 (contribución del año 1958 al 1999) se constituye a partir de modelado numérico de alta resolución de atmósfera, nivel del mar y oleaje cubriendo todo el entorno litoral español.

Las series WANA (contribución del año 2000 a la actualidad) proceden del sistema de predicción del estado de la mar que Puertos del Estado ha desarrollado

en colaboración con la Agencia Estatal de Meteorología. No obstante, los datos WANA (ver Puertos del Estado (b)) no son datos de predicción sino datos de diagnóstico o análisis. Esto supone que para cada instante el modelo proporciona campos de viento y presión consistentes con la evolución anterior de los parámetros modelados y consistentes con las observaciones realizadas. Para generar los campos de oleaje se ha utilizado el modelo numérico WAM. Dicha aplicación es un modelo espectral de tercera generación que resuelve la ecuación de balance de energía sin establecer ninguna hipótesis *a priori* sobre la forma del espectro de oleaje. Los datos se han generado con una cadencia horaria. En el área mediterránea se ha generado datos en una malla de nodos de separación variable con una resolución de 15' de latitud x 15' de longitud (unos 25 Km x 25 Km) para el borde Este de la malla y de 7.5' de latitud x 7.5' de longitud (aproximadamente 12.5Km x 12.5Km) para el resto del área modelada. El modelo WAM utilizado para generar estos datos incluye efectos de refracción y asomeramiento. No obstante, dada la resolución del modelo, se pueden considerar despreciables los efectos del fondo. Por tanto, para uso práctico los datos de oleaje deben de interpretarse siempre como datos en aguas abiertas a profundidades indefinidas.

4.2. Parámetros disponibles

La serie de datos SIMAR consta de diferentes parámetros de viento y oleaje (altura significativa espectral, periodo de pico espectral, periodo medio espectral, dirección media de procedencia del oleaje y dirección de mar de viento, periodo medio y dirección de mar de fondo) con frecuencia horaria. Los datos en que nos centraremos para la realización de esta tesina son los de altura significativa, que se define como la media aritmética del tercio superior de datos del registro. Este es el estadístico que se utiliza habitualmente en los estudios de clima marítimo y en el diseño de obras marítimas.

Capítulo 5

Caracterización del oleaje

A continuación se describen las series de datos del oleaje correspondientes al nodo SIMAR 2122142 indicado en la figura 4.1, cerca de Sant Feliu de Guíxols, durante el periodo temporal Enero 1958-Noviembre 2014, con datos cada 3 horas. Se incluyen distribuciones conjuntas de altura y periodo, distribuciones conjuntas de altura y dirección de oleaje y rosas de oleaje. Para describir el clima extremo se ajusta la distribución Generalizada de Pareto y distribuciones de Weibull por cada dirección.

5.1. Distribución conjunta de periodo de pico y altura significativa

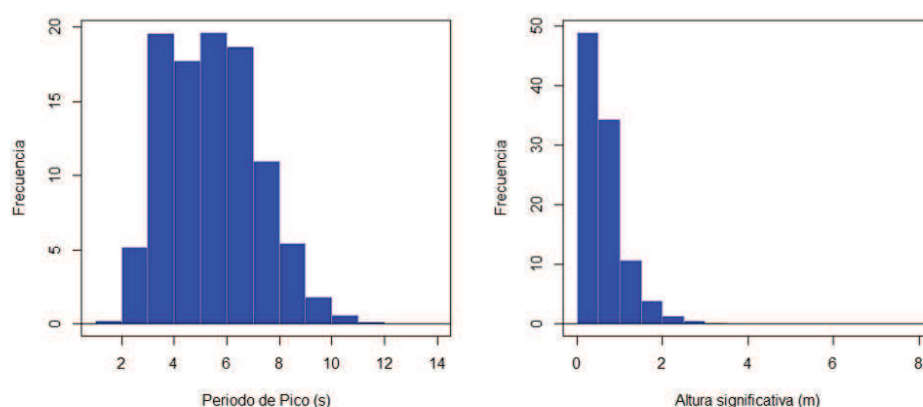


Figura 5.1: Histogramas de Tp y Hs.

Hs(m)/Tp(s)	<2	3	4	5	6	7	8	9	10	>10	Total (%)
<0,5	0,221	4,930	16,046	11,638	9,786	5,120	0,970	0,213	0,049	0,013	48,986
1	0,001	0,262	3,497	5,743	8,327	10,006	4,830	1,324	0,213	0,052	34,255
1,5	0,000	0,002	0,054	0,366	1,484	2,976	3,550	1,869	0,352	0,046	10,698
2	0,000	0,000	0,000	0,009	7,335	0,520	1,214	1,324	0,498	0,121	3,760
2,5	0,000	0,000	0,000	0,002	0,001	0,052	0,329	0,490	0,366	0,138	1,377
3	0,000	0,000	0,000	0,000	0,001	0,005	0,061	0,161	0,190	0,128	0,545
3,5	0,000	0,000	0,000	0,000	0,000	0,000	0,005	0,039	0,095	0,088	0,228
4	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,008	0,034	0,052	0,095
4,5	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,007	0,025	0,032
5	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,014	0,014
>5	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,009	0,009
Total (%)	0,222	5,194	19,596	17,758	19,671	18,678	10,960	5,429	1,804	0,687	100,000

Figura 5.2: Tabla Periodo de Pico - Altura Significante.

La figura 5.1 muestra los histogramas de periodo de pico y altura significativa. Se observa que los valores más comunes para el periodo de pico estan entre los 3 y 7 segundos y para la altura significativa entre 0 y 1 metro. La altura significativa caracteriza de modo principal la severidad de un temporal. No obstante, la acción de un temporal sobre una estructura también depende del periodo del oleaje. Por

ello, una vez seleccionados los picos de temporal se establece una relación empírica entre el periodo de pico y la altura significativa del oleaje ajustando por mínimos cuadrados una relación del tipo: $E(T_p) = aH_s^c$.

$E(T_p)$ es el valor esperado del periodo de pico para el pico de un temporal de altura significativa H_s . Por tanto, es de interés el estudio conjunto de estas dos variables. La tabla 5.2 muestra que los valores más frecuentes son de 0.5 a 1.5 metros y de 3 a 7 segundos, para la altura de ola significativa y periodo de pico respectivamente. Para mostrar los valores de periodo de pico y altura significativa de una forma más visual se utiliza el diagrama 5.3 donde el color más oscuro representa más densidad de datos. En el diagrama de dispersión de altura significativa - periodo de pico se ve que existe una gran dispersión entre la altura de ola significativa y el periodo de pico del oleaje. En las alturas de ola pequeñas se ha registrado periodos de pico entre 0 y 12 segundos, mientras que al ir aumentando la altura de ola el rango de valores se va reduciendo, para 2 metros de altura de ola los periodos están entre 4 y 12 segundos y para más de 6 metros están entre 10 y 14. Esta relación entre altura de ola significativa y periodo de pico es típica de las fuentes de oleaje.

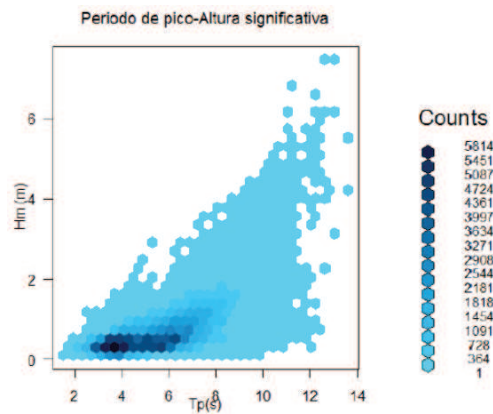


Figura 5.3: Distribución conjunta de H_s - T_p .

5.2. Distribución conjunta de dirección y altura significativa

En la rosa de oleaje de la figura 5.4 se representan la altura y dirección del oleaje asociadas a su probabilidad de ocurrencia. En la figura 5.4 los colores indican la altura de ola y la longitud es proporcional a la frecuencia de cada sector. Tal como se puede observar tanto en la rosa de oleaje como en la tabla de contingencia, los oleajes más frecuentes son los de altura de ola hasta 1.5 metros. En cuanto a la distribución sectorial del oleaje, se puede ver que los sectores N, NNE y SSW son los que presentan oleajes con mayor altura de ola, llegando a alcanzar los 5m.

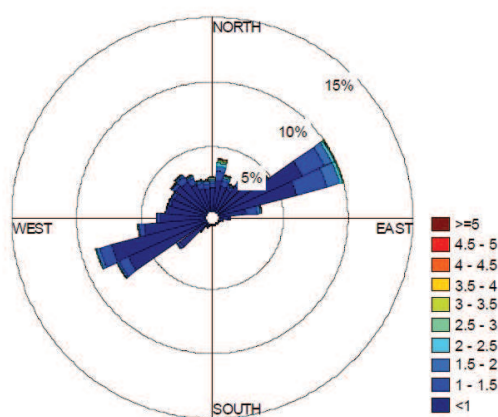


Figura 5.4: Rosa de Oleaje.

CAPÍTULO 5. CARACTERIZACIÓN DEL OLEAJE

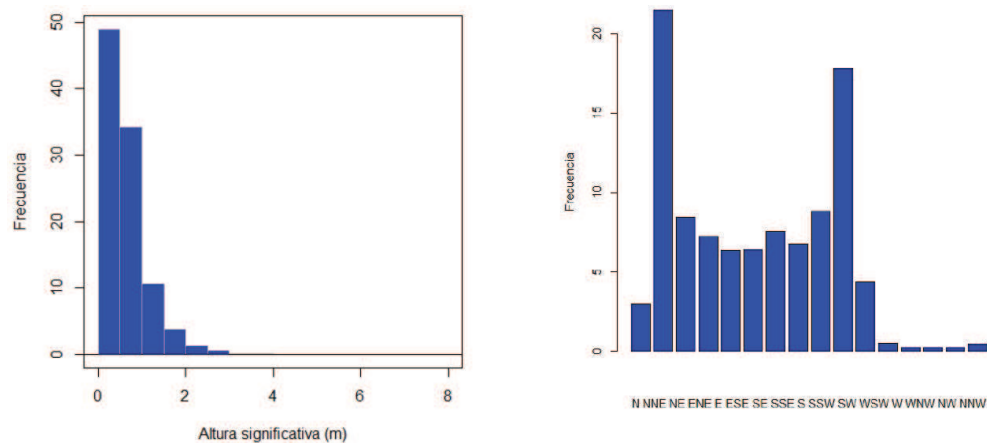


Figura 5.5: Histogramas de Altura Significante y Dirección.

Dirección/Hs(m)	<0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	>5,5	Total (%)
N	1,210	1,138	0,493	0,140	0,019	0,002	0,005	0,003	0,000	0,000	0,000	3,009
NNE	7,010	8,553	4,043	1,396	0,363	0,099	0,026	0,002	0,000	0,000	0,000	21,493
NE	4,653	2,676	0,728	0,278	0,088	0,032	0,014	0,001	0,001	0,000	0,002	8,474
ENE	3,182	1,987	0,922	0,511	0,307	0,171	0,090	0,042	0,020	0,014	0,011	7,257
E	2,590	2,015	0,820	0,443	0,254	0,119	0,061	0,038	0,014	0,005	0,008	6,368
ESE	2,579	2,461	0,800	0,324	0,158	0,075	0,023	0,008	0,002	0,001	0,000	6,431
SE	3,729	2,939	0,636	0,181	0,055	0,013	0,007	0,003	0,001	0,000	0,000	7,565
SSE	4,286	2,101	0,300	0,056	0,020	0,004	0,002	0,000	0,000	0,000	0,000	6,768
S	5,335	2,983	0,372	0,097	0,029	0,006	0,001	0,000	0,000	0,000	0,000	8,824
SSW	9,966	6,079	1,362	0,307	0,087	0,024	0,003	0,000	0,000	0,000	0,000	17,827
SW	3,043	1,081	0,191	0,028	0,004	0,001	0,000	0,000	0,000	0,000	0,000	4,347
WSW	0,424	0,059	0,007	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,490
W	0,210	0,029	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,242
WNW	0,172	0,025	0,002	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,200
NW	0,192	0,040	0,007	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,239
NNW	0,357	0,096	0,013	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,467
Total	48,938	34,263	10,698	3,765	1,384	0,544	0,231	0,097	0,037	0,021	0,021	100,000

Figura 5.6: Tabla Altura Significante - Dirección.

5.3. Distribución de Weibull por direcciones

Se dispone de una serie con datos cada 3 horas. Para el estudio del comportamiento extremal del oleaje se seleccionarán los datos de altura de ola superiores a 2 metros, a partir de este umbral se considera que ha habido un temporal. La altura de ola que caracteriza cada temporal es el valor máximo de todos los datos contiguos que superan el umbral de 2 metros. Para asegurar que los temporales seleccionados son independientes, se ha comprobado que la distancia mínima entre el final de un temporal y el inicio del siguiente sea superior a 72 horas. Aplicando este método se consigue pasar de una serie continua a un conjunto de valores de altura de ola que caracterizan los temporales registrados por cada nodo SIMAR. El gráfico 5.7 muestra una serie hipotética de datos de altura significativa y se ilustra como se realiza la selección de los valores que representan el comportamiento extremal de la serie de valores de altura significativa.

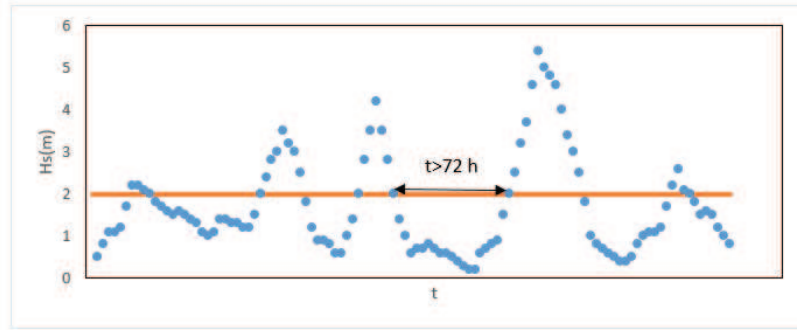


Figura 5.7: Serie de datos de altura de ola.

En los estudios de clima marítimo es habitual suponer que la altura de ola máxima sigue una distribución Weibull. Por este motivo, se ha utilizado esta distribución que tiene como función de probabilidad acumulada la siguiente:

$$F(x) = 1 - \exp\left(-\frac{x - A}{B}\right)^C. \quad (5.1)$$

Donde A es el parámetro de centrado, B es el parámetro de escala y C es el parámetro de forma. Se ajustan las distribuciones, por direcciones, a los datos de altura de ola y se observa que la dirección de los temporales con mayor altura de ola es el este.

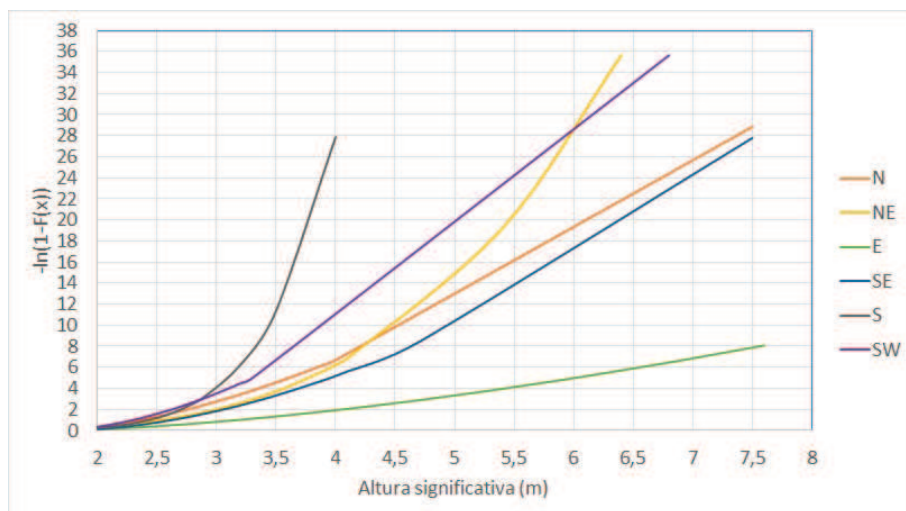


Figura 5.8: Distribuciones de Weibull.

Los parámetros de las distribuciones de Weibull ajustadas, según dirección son los siguientes:

Dirección	A	B	C
N	0,72	1,64	1,61
NE	1,96	0,57	3,28
E	1,55	1,63	1,55
SE	0,92	1,7	1,81
S	2,44	0	6,71
SW	0,66	1,64	1,74

Figura 5.9: Parámetros de las distribuciones de Weibull por direcciones.

Las distribuciones de Weibull se han ajustado utilizando Excel. Primero se ha utilizado una transformación logarítmica y se ha fijado el valor del parámetro C. De este modo encontramos A (punto de intersección de la recta con las abscisas) y B (pendiente de la recta). El parámetro C se ha obtenido maximizando el coeficiente R^2 , pues el objetivo es mejorar el ajuste de la recta de regresión.

5.4. Distribución Generalizada de Pareto

El teorema de Pickands (Pickands (1975)) asegura que, dado un umbral suficientemente alto, los excesos de altura de ola sobre este umbral siguen una distribución generalizada de Pareto. Por tanto, para describir el clima extremo se ajusta, a los valores de altura de ola superiores a 2 metros, la distribución Generalizada de Pareto cuya expresión de probabilidad acumulada es la 5.2. Para ajustar esta distribución se ha utilizado el método de máxima verosimilitud.

$$F(x) = 1 - \exp \left(\frac{1}{\kappa} \log \left(\frac{1 - \kappa(x - \xi)}{\alpha} \right) \right). \quad (5.2)$$

Donde ξ es el parámetro de localización, α es el parámetro de escala y κ es el parámetro de forma. Ajustar los datos a una distribución teórica permite calcular la probabilidad con la que se puede presentar un temporal de una cierta altura de ola. Dada la serie de datos de altura de ola, ajustamos la distribución a los datos superiores a 2 metros y se obtiene los siguientes valores de los parámetros: $\xi = 1.95$ $\kappa = -0.083$ $\alpha = 0.512$. En la figura 5.10 se muestra la distribución de probabilidad acumulada y en la figura 5.11, la función de densidad.

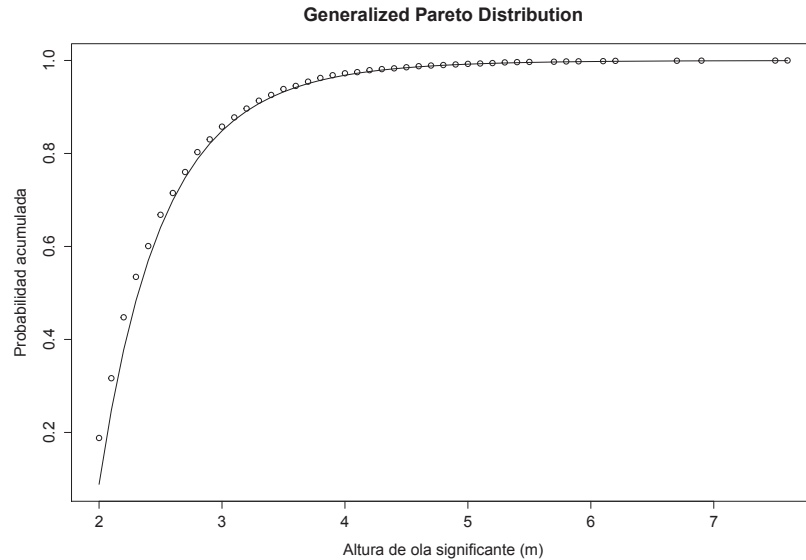


Figura 5.10: Función de probabilidad acumulada de la distribución generalizada de Pareto.

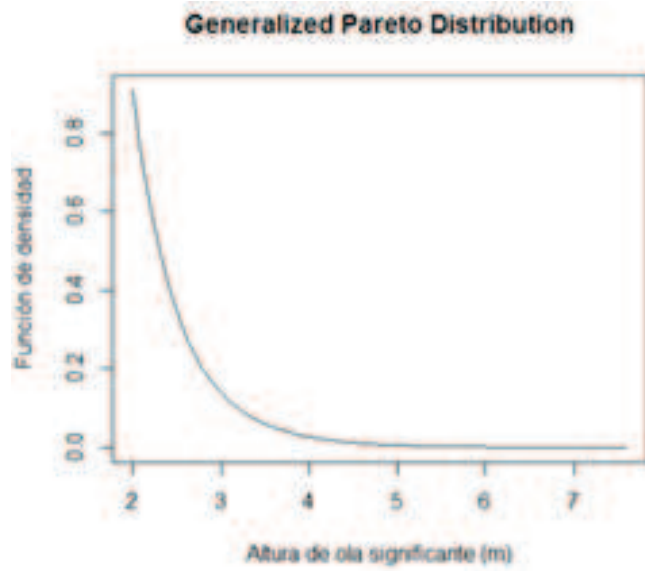


Figura 5.11: Función de densidad de la distribución generalizada de Pareto.

Capítulo 6

Datos composicionales

6.1. Definición de categorías

Para el estudio del comportamiento extremal del oleaje se deben seleccionar aquellos sucesos que pueden ser considerados extremos. Para asegurar independencia entre estos sucesos se establece un criterio de selección. Se dispone de datos de altura de ola cada hora y se quiere estudiar la evolución de la altura de ola máxima de cada temporal. Para encontrar los valores de los picos de cada temporal se selecciona los valores de altura de ola que superan el umbral de 2 metros y se supone que el tiempo mínimo que transcurre entre dos temporales independientes es de 3 días. Un temporal queda representado por el pico o valor máximo de altura de ola. El método de selección de temporales descrito se conoce como POT (Peak Over Threshold) (figura 6.1).

Como el número de temporales de cada año es bajo se plantea un enfoque alternativo, se transformará la variable altura de ola en una variable categórica. Se definen cinco categorías en función de la altura de ola de forma que cada temporal pertenecerá a una de estas categorías: débil, moderadamente débil, moderado, moderadamente intenso o intenso. La variable transformada cuenta el número de temporales en cada categoría y sigue una distribución multinomial (expresión 6.1). El objetivo es estimar los parámetros de esta distribución que son las probabilidades de cada categoría.

$$Pr[X_1 = x_1, \dots, X_5 = x_5] = \frac{n!}{x_1! \dots x_5!} p_1^{x_1} \dots p_5^{x_5}. \quad (6.1)$$

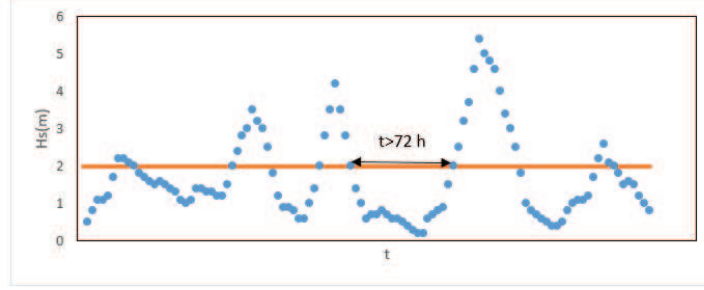


Figura 6.1: Serie de datos de altura de ola. Separación de 72h entre temporales independientes.

Se definirán los intervalos de forma que el incremento relativo sea constante, no el absoluto. Para definir las clases según este criterio se hace una transformación logarítmica de coordenadas $z = \ln(Y)$ y se obtiene el intervalo considerando el rango de altura de ola, que va de 2 a 7.6 m para el caso del nodo SIMAR 2122142:

$$l = \frac{\ln(H_f) - \ln(H_o)}{k}. \quad (6.2)$$

Se definen las categorías en función de H haciendo la transformación inversa $Y = \exp(Z)$:

$$H_i = (\exp(Z_o + (i - 1) \cdot l), \exp(Z_o + i \cdot l)). \quad (6.3)$$

Finalmente los rangos quedan definidos de la siguiente forma:

$$H_1 = [2, 2.61), H_2 = [2.61, 3.41), H_3 = [3.41, 4.46), H_4 = [4.46, 5.82), H_5 = [5.82, \infty)$$

Con las categorías definidas ya se pueden clasificar los temporales según el año en que se produjeron y la clase a la que pertenecen. De esta forma, cada año queda caracterizado por los 5 valores que corresponden al número de temporales registrados en cada una de las categorías. El resultado es una matriz como la que se presenta a continuación en la tabla 6.2. Con esta transformación se ha pasado de analizar una variable univariante a una multivariante y los datos transformados son datos composicionales. Se observa que en las categorías que corresponden a temporales más intensos se han registrado pocos temporales.

	Débil	Mod. Débil	Moderado	Mod. intenso	Intenso
1958	6	3	1	0	0
1959	4	7	0	0	1
1960	3	3	1	1	0
1961	3	1	2	0	0
1962	8	3	1	1	0
1963	5	4	2	0	0
1964	4	2	0	0	1
1965	4	4	0	0	0
1966	3	3	0	0	0
1967	3	1	3	0	0
1968	12	1	0	0	0
1969	4	4	2	0	0
1970	3	1	1	1	0
1971	5	2	3	0	0
1972	5	2	1	1	1
1973	7	0	2	0	0
1974	2	1	0	1	0
1975	3	1	1	1	0
1976	6	0	2	0	0
1977	4	1	0	1	0
1978	1	3	1	1	0
1979	7	3	0	0	0
1980	5	3	1	1	0
1981	5	1	0	0	0
1982	5	1	2	2	0
1983	8	2	2	0	0
1984	7	1	4	0	0
1985	5	2	0	1	0
1986	5	2	0	1	0
1987	5	2	1	0	0
1988	3	2	2	0	0
1989	7	2	1	0	0
1990	5	3	0	0	0

Figura 6.2: Datos clasificados por categoría y año.

6.2. Análisis composicional de datos

Como se ha introducido en el apartado anterior, se han definido 5 categorías y se ha contado el número de temporales en cada categoría y cada año. Se clausura cada fila de la matriz anterior para que la suma de las componentes del vector sea constante. El vector de proporciones tiene todas las componentes positivas y la suma es constante por tanto tiene las características de los datos composicionales y pertenece al espacio símplex. A continuación se hace un resumen del análisis de datos composicionales, para más información ver Pawlowsky-Glahn *et al.* (2015).

Dada una matriz de datos composicionales formada por vectores composicionales con D partes,

$$X = x_i = [x_{i1}, \dots, x_{iD}] \quad (6.4)$$

se puede calcular la media composicional de X , el centro, de la siguiente forma:

$$g_i = \left(\prod_{j=1}^D x_{ij} \right)^{1/n}. \quad (6.5)$$

Dada la matriz de datos $X \in S^D$, la varianza total tiene la expresión:

$$TotVar[X] = \frac{1}{n} \sum_{i=1}^n d_a^2(x_i, g). \quad (6.6)$$

Donde d_a es la distancia de Aitchison definida en el espacio símplex:

$$d_a(x, y) = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2}. \quad (6.7)$$

y la matriz de variación es:

$$T = \begin{pmatrix} Var \left(\ln \frac{x_1}{x_1} \right) & Var \left(\ln \frac{x_1}{x_2} \right) & \dots & Var \left(\ln \frac{x_1}{x_D} \right) \\ Var \left(\ln \frac{x_2}{x_1} \right) & Var \left(\ln \frac{x_2}{x_2} \right) & \dots & Var \left(\ln \frac{x_2}{x_D} \right) \\ \vdots & \vdots & \ddots & \vdots \\ Var \left(\ln \frac{x_D}{x_1} \right) & Var \left(\ln \frac{x_D}{x_2} \right) & \dots & Var \left(\ln \frac{x_D}{x_D} \right) \end{pmatrix}.$$

Las componentes de la matriz de variación representan la asociación entre dos variables, si el valor es grande significa que las dos variables correspondientes están disociadas. En este caso:

$$TotVar = 1,6.$$

$$T = \begin{pmatrix} 0 & 0,14 & 0,41 & 0,69 & 0,23 \\ 0,14 & 0 & 0,67 & 1,04 & 0,47 \\ 0,41 & 0,67 & 0 & 0,84 & 0,89 \\ 0,69 & 1,04 & 0,84 & 0 & 0,60 \\ 0,23 & 0,47 & 0,89 & 0,60 & 0 \end{pmatrix}.$$

El número de temporales en las cinco categorías sigue una distribución multinomial. Para poder estudiar esta variable aleatoria multivariante necesitamos transformar los datos del espacio símplex al real. Para pasar el vector del espacio símplex y sujeto a una restricción a un vector real de dimensión D-1, sin restricciones, se utiliza la transformación ilr (isometric log-ratio):

$$x^* = ilr(x) = clr(x) \cdot \Psi', x = C(\exp(x^* \Psi)). \quad (6.8)$$

Donde Ψ es la matriz clr de los vectores de la base:

$$\Psi = \begin{pmatrix} clr(e_1) \\ clr(e_2) \\ \dots \\ clr(e_{D-1}) \end{pmatrix}, \Psi \Psi' = I_{D-1}.$$

La transformación clr de un vector se define:

$$clr(x) = \left(\ln \frac{x_1}{g(x)}, \dots, \ln \frac{x_D}{g(x)} \right). \quad (6.9)$$

Para poder hacer la transformación primero se construye una base ortonormal del símplex asociada a una partición binaria secuencial y después se aplica la transformación ilr que transforma el vector perteneciente al símplex y sujeto a una restricción en un vector real de coordenadas de dimensión D-1 sin restricciones. Las coordenadas de este vector se pueden entender como balances entre grupos de partes de la partición. Nos interesa estudiar la evolución del balance entre los temporales poco intensos y moderados frente los extremos, por ese motivo en la primera partición separamos las dos categorías más intensas del resto.

Definimos la siguiente partición binaria secuencial de la composición de 5 partes:

$$\begin{pmatrix} 1 & 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \end{pmatrix}.$$

1. Separamos los grupos de partes {débil, moderadamente débil, moderado} y {moderadamente intenso, intenso}.
2. En el grupo {mod. intenso, intenso} se separa {moderadamente intenso} y {intenso}.
3. En el grupo { débil, mod. débil, moderado} se separa {débil, moderadamente débil} y {moderado}.
4. Finalmente se separa {débil} y {moderadamente débil}.

Cada partición tiene asociado un vector de la base ortonormal. La expresión de los vectores de la base es la siguiente:

$$\begin{aligned} e_1 &= C \left[\exp \left(\frac{\sqrt{2}}{\sqrt{3 \cdot 5}}, \frac{\sqrt{2}}{\sqrt{3 \cdot 5}}, \frac{\sqrt{2}}{\sqrt{3 \cdot 5}}, \frac{-\sqrt{3}}{\sqrt{2 \cdot 5}}, \frac{-\sqrt{3}}{\sqrt{2 \cdot 5}} \right) \right]. \\ e_2 &= C \left[0, 0, 0, \exp \left(\frac{1}{\sqrt{1 \cdot 2}}, \frac{-1}{\sqrt{1 \cdot 2}} \right) \right]. \\ e_3 &= C \left[\exp \left(\frac{1}{\sqrt{2 \cdot 3}}, \frac{1}{\sqrt{2 \cdot 3}}, \frac{\sqrt{-2}}{\sqrt{1 \cdot 3}}, 0, 0 \right) \right]. \\ e_4 &= C \left[\exp \left(\frac{1}{\sqrt{1 \cdot 2}}, \frac{-1}{\sqrt{1 \cdot 2}}, 0, 0, 0 \right) \right]. \end{aligned}$$

La matrix de contraste de la base es:

$$\Psi = \begin{pmatrix} 0,37 & 0 & 0,41 & 0,71 \\ 0,37 & 0 & 0,41 & -0,71 \\ 0,37 & 0 & -0,82 & 0 \\ -0,55 & 0,71 & 0 & 0 \\ -0,55 & -0,71 & 0 & 0 \end{pmatrix}.$$

La coordenada c_i es la proyección en la dirección del vector de la base e_i . La expresión del vector de coordenadas asociado a la base ortonormal definida anteriormente son log cocientes, la expresión de los cuales es:

$$c_1 = \ln \left[\frac{(x_1 x_2 x_3) \sqrt{2/15}}{(x_4 x_5) \sqrt{3/10}} \right].$$

$$c_2 = \ln \left[\frac{(x_4) \sqrt{1/2}}{(x_5) \sqrt{1/2}} \right].$$

$$c_3 = \ln \left[\frac{(x_1 x_2) \sqrt{1/6}}{(x_3) \sqrt{2/3}} \right].$$

$$c_4 = \ln \left[\frac{(x_1) \sqrt{1/2}}{(x_2) \sqrt{1/2}} \right].$$

La transformación *ilr* nos ha permitido pasar del espacio símplex al real y por tanto se podrá aplicar los métodos estadísticos multivariantes de espacios reales. Para volver al espacio símplex se aplicará la transformación *ilr* inversa.

6.3. Dendrograma de la partición

El dendrograma composicional es un tipo de representación gráfica que nos ayuda a apreciar claramente las relaciones entre las categorías y nos permite visualizar simultáneamente la base de balances elegida, las medias de los balances y su varianza. Los valores medios de los balances están representados en el dendrograma por los puntos de enlace de las líneas horizontales con las verticales. En la figura 6.3 se aprecia que la mayor parte de varianza está asociada a los log-cocientes de los temporales intensos y moderadamente intensos sobre el resto.

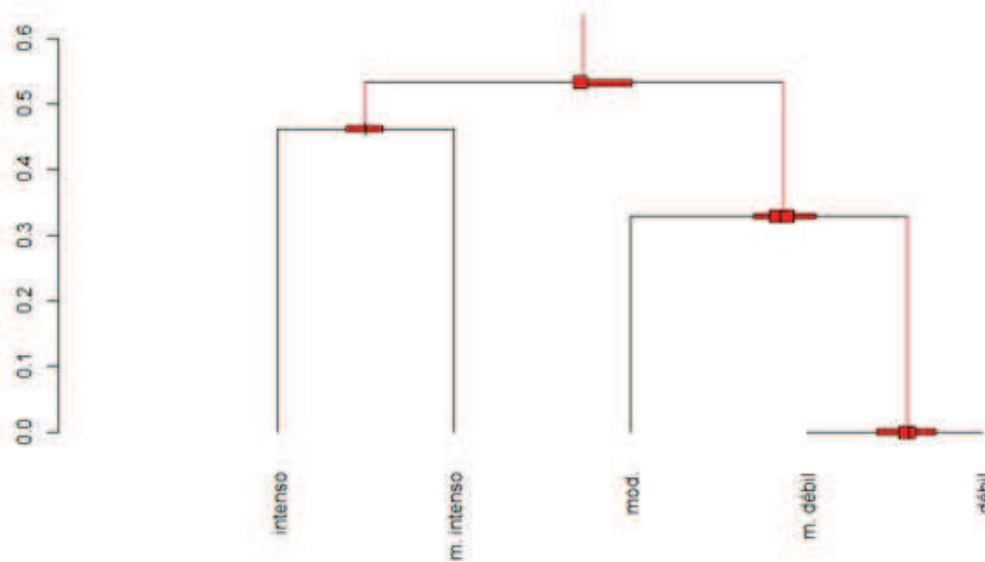


Figura 6.3: Dendrograma donde se representa la media y varianza de las coordenadas. Los valores medios de los balances están representados en el dendrograma por los puntos de enlace de las líneas horizontales con las verticales. La varianza queda representada por los diagramas de caja.

Capítulo 7

Estimación de los parámetros

7.1. Estimador de máxima verosimilitud de la distribución multinomial

Se necesita estimar los parámetros de la distribución multinomial que corresponden a las probabilidades de ocurrencia de las categorías. Una forma de hacerlo sería utilizando el estimador de máxima verosimilitud de esta distribución. Sin embargo, utilizar este método clásico no es lo más adecuado ya que el número de temporales extremos es bajo y en algunos casos nulo. El inconveniente que tiene el estimador puntual de máxima verosimilitud de una probabilidad multinomial es que es no regular. Si el número de sucesos observados en una categoría es nulo la probabilidad estimada de esta categoría será cero, esto quiere decir que los temporales de esta categoría serían un suceso imposible, cosa que no es cierta. Además los métodos frecuentistas suponen que los parámetros son fijos y desconocidos. En nuestro caso se dispone de pocos datos y por tanto la incertidumbre es alta. En estas condiciones, puede resultar conveniente utilizar la estadística bayesiana.

El estimador de máxima verosimilitud se utilizará como aproximación inicial para generar una muestra del vector de proporciones de la distribución multinomial con métodos de Montecarlo (MCMC). La función de verosimilitud de los

parámetros desconocidos es:

$$L[p_1, \dots, p_5 | X_1 = x_1, \dots, X_5 = x_5] = \frac{n!}{x_1! \dots x_5!} p_1^{x_1} \dots p_5^{x_5}. \quad (7.1)$$

La función de log-verosimilitud es:

$$\ln(L) = \ln \left(\frac{n!}{x_1! \dots x_5!} \right) + \sum_{j=1}^k x_j \ln p_j. \quad (7.2)$$

La estimación de máxima verosimilitud son los valores que maximizan la función anterior y que cumplen la restricción:

$$p_1 + p_2 + p_3 + \dots + p_k = 1.$$

Dado $X_i = x_i$, el número de sucesos observados en la categoría i , y n el número total de sucesos, el estimador de máxima verosimilitud es $p_i = x_i/n$.

7.2. Estadística bayesiana

Para detectar si ha habido un cambio en el clima extremal primero necesitamos estimar el vector de proporciones $\theta = (c_1, c_2, c_3, c_4)$. Mediante métodos bayesianos y con el algoritmo de Hamilton Monte Carlo (ver Robert & Casella (2004)) generaremos una muestra simulada del vector de proporciones y podremos analizar los resultados. Existe un gran número de libros que se pueden utilizar como introducción a los métodos bayesianos como: Gelman *et al.* (2014), Albert (2007) y Carlin & Louis (2009). La estimación bayesiana se basa en suponer que los parámetros desconocidos no tienen un valor fijo sino que son variables aleatorias que siguen una densidad de probabilidad que modela la incertidumbre acerca de su valor. Siguiendo el teorema de Bayes a partir de la distribución de probabilidad de los datos *a priori*, dado el modelo y los datos, la probabilidad conjunta de los parámetros del modelo condicionada a los datos es:

$$p(\theta|x) = p(\theta)p(x|\theta). \quad (7.3)$$

7.2.1. Modelo en el espacio simplex

Se quiere estimar los parámetros de una multinomial. Si se utiliza como función *a priori* la distribución de Dirichlet con parámetros α :

$$p(p_1, \dots, p_5) = \frac{1}{B(\alpha)} \prod_{i=1}^5 p_i^{\alpha_i - 1}. \quad (7.4)$$

Donde $B(\alpha)$ es la función beta. Y la verosimilitud es una multinomial:

$$L[X_1 = x_1, \dots, X_5 = x_5 | p_1, \dots, p_5] = \frac{n!}{x_1! \dots x_5!} p_1^{x_1} \dots p_5^{x_5}. \quad (7.5)$$

La distribución *posteriori* es también una Dirichlet con parámetros $(\alpha_1 + x_1, \dots, \alpha_5 + x_5)$. La distribución Dirichlet es el *priori* conjugado de la distribución multinomial ya que *priori* y *posteriori* siguen el mismo tipo de distribución.

7.2.2. Modelo en el espacio real

La distribución *a priori* se basa en la información que se posee y expresa el grado de creencia que se tiene antes de estudiar los datos. Como en este caso se dispone de poca información *a priori* la distribución que se utilizará es una normal multivariante objetiva que tiene la expresión:

$$f_{\text{priori}}(\theta_1, \dots, \theta_n) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\theta - \mu)^T \Sigma^{-1}(\theta - \mu)\right). \quad (7.6)$$

Para que la función *a priori* sea objetiva las varianzas de la matriz de covarianzas tienen que ser grandes. De esta forma la función *a posteriori* se parecerá a la verosimilitud de los datos y la mayor parte de la información que se aportará al modelo será la de los datos. La función de verosimilitud introduce la información muestral que contienen los datos a nuestros conocimientos previos, es la probabilidad que los datos sean generados por el modelo dados unos valores de los parámetros. La función de verosimilitud de las coordenadas del modelo para los datos \mathbf{X} es una multinomial. Para trabajar en coordenadas se ha escrito las probabilidades en función de las coordenadas utilizando la transformación ilr inversa. Considerando que cada fila de la matriz es independiente la verosimilitud es el producto de la ecuación 7.7 aplicada a los datos de cada año. Como se explica en Pawlowsky-Glahn *et al.* (2015) la función de probabilidad multinomial,

en coordenadas, que resulta después del cambio de variable tiene la siguiente expresión:

$$f[x|\theta] = \frac{N!}{\prod_{i=1}^D x_i!} \exp \left[N \ln(C(\theta)) + \sum_1^{D-1} \theta_j \left(\sum_1^D x_i \psi_{ji} \right) \right]. \quad (7.7)$$

Multiplicando la distribución *a priori* por la verosimilitud se obtiene la función de densidad *posteriori* que contiene la información actualizada tras el análisis de los datos y tiene la siguiente expresión:

$$f_{posteriori}(\theta|x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\theta - \mu)^T \Sigma^{-1} (\theta - \mu) \right) \frac{N!}{\prod_{i=1}^D x_i!} \exp \left[N \ln(C(\theta)) + \sum_1^{D-1} \theta_j \left(\sum_1^D x_i \psi_{ji} \right) \right].$$

La distribución conjunta *a posteriori* resultante tiene una expresión complicada, por tanto será necesario utilizar los métodos Monte Carlo Markov Chain (ver Robert & Casella (2004)) para simular una muestra de la distribución *a posteriori* a partir de la cual se podrá realizar cualquier inferencia sin necesidad de integrar la función *posteriori*. El método de muestreo que se utilizará en este caso es el de Hamilton Monte Carlo (ver Gelman *et al.* (2014)). La muestra se obtiene en coordenadas, para pasar del espacio real al simplex se utiliza la transformación ilr inversa.

7.3. Método de Hamilton Monte Carlo

También se conoce como Monte Carlo híbrido ya que combina métodos de muestreo deterministas con MCMC. El método de Hamilton MC converge más rápido a la distribución objetivo que otros como el de Gibbs (ver Kruschke (2010)) ya que los pasos no son aleatorios. Para cada variable θ de la distribución *posteriori* el método de Hamilton MC añade una variable ϕ que controla los pasos de las variables θ . Esta variable auxiliar permite al algoritmo moverse más rápidamente por el espacio de los parámetros. HMC necesita el gradiente de la log-densidad *posteriori*. Si θ tiene D componentes este gradiente es:

$$\frac{d \ln p(\theta|y)}{d\theta} = \left(\frac{d \ln p(\theta|y)}{d\theta_1}, \dots, \frac{d \ln p(\theta|y)}{d\theta_D} \right).$$

Cada iteración tiene tres pasos:

1. La iteración se inicia mediante la actualización al azar de ϕ a partir de su distribución *posteriori* que es la misma que su distribución *priori*.

2. Actualización simultánea de θ y ϕ . Esta actualización consiste en L saltos, cada uno escalado por un factor ϵ , como se explica a continuación. Se repiten los siguientes pasos L veces.

a) Se usa el gradiente de la log-densidad *a posteriori* de θ para actualizar ϕ :

$$\phi \leftarrow \phi + \frac{1}{2}\epsilon \frac{d \ln p(\theta|y)}{d\theta}. \quad (7.8)$$

b) Se usa el vector ϕ para actualizar la posición del vector θ .

$$\theta \leftarrow \theta + \epsilon M^{-1} \phi. \quad (7.9)$$

M es la matriz de covarianzas de la distribución normal que sigue la variable auxiliar ϕ .

c) Se vuelve a usar el gradiente de la log-densidad *a posteriori* de θ para actualizar ϕ :

$$\phi \leftarrow \phi + \frac{1}{2}\epsilon \frac{d \ln p(\theta|y)}{d\theta}. \quad (7.10)$$

3. Siendo θ^{t-1} y ϕ^{t-1} los valores de los parámetros y de la variable auxiliar en el primero de los L pasos y θ^* , ϕ^* los valores después de los L pasos, los valores actualizados se aceptan o rechazan según:

$$r = \frac{p(\theta^*|y)p(\phi^*)}{p(\theta^{t-1}|y)p(\phi^{t-1})}. \quad (7.11)$$

4. El criterio de aceptación es el siguiente:

$$\theta^t = \begin{cases} \theta^* & \text{con probabilidad } \min(r, 1) \\ \theta^{t-1} & \text{en el resto de casos} \end{cases}. \quad (7.12)$$

Al igual que con cualquier otro algoritmo MCMC, repetimos estas iteraciones hasta la convergencia y el tamaño efectivo de la muestra sea lo suficientemente grande.

Capítulo 8

Regionalización de la altura de ola

El objetivo de este capítulo es agrupar los nodos SIMAR en clústers considerando los valores de las probabilidades de cada categoría y teniendo en cuenta la situación espacial de cada nodo, como se hace en Ortego *et al.* (2011) para datos pluviométricos. La dependencia espacial de los datos composicionales se caracteriza mediante los variogramas (ver Ortego *et al.* (2011) y Pawlowsky-Glahn & Olea (2004)). Se utiliza un método clúster jerárquico para agrupar los nodos con mayor correlación (ver Everitt (1993)). Para evaluar esta correlación se define la distancia Mahalanobis entre nodos utilizando los variogramas de los log-ratios de cada par de variables, que caracterizan la dependencia espacial de los datos.

8.1. Variogramas.

Los variogramas de los log-ratios de cada par de variables (i, j) de los datos composicionales se calculan con la siguiente expresión:

$$\gamma_{ij}(h) = \frac{1}{2N(h)} \sum_{x,z \in N(h)} \left(\ln \frac{p_{xi}}{p_{xj}} - \ln \frac{p_{zi}}{p_{zj}} \right)^2. \quad (8.1)$$

Para ajustar un modelo exponencial a los variogramas empíricos se utiliza un correlograma paramétrico $\rho(h, \theta)$ que depende del parámetro θ y se optimiza la

diferencia logarítmica entre los variogramas empíricos y el modelo ajustado:

$$\Psi(h, \theta, A, B) = A\delta(h \neq 0) + B(1 - \rho(h, \theta)). \quad (8.2)$$

De acuerdo con Tolosana *et al.* (2011), se impone la restricción que las dos matrices A (*nugget effect*) y B (*variogram sill*) son negativas semidefinidas, para asegurar que el modelo $\Psi(h, \theta, A, B)$ sea compatible con la función de covarianza para las coordenadas ilr :

$$C(h, \theta, \dots) = -\frac{1}{2}VAV^t\delta(h \neq 0) + VBV^t(1 - \rho(h, \theta)). \quad (8.3)$$

Para $h \neq 0$ esta expresión describe la dependencia espacial entre las variables consideradas, mientras que para $h = 0$ refleja la covarianza clásica entre ellas. Esta matriz se utilizará en la ecuación 8.6 para calcular la distancia de Mahalanobis y de esta forma se tendrá en cuenta la dependencia espacial. Los variogramas de los log-ratios de cada par de variables y los modelos exponenciales, sin *nugget*, ajustados son los de la figura 8.1.

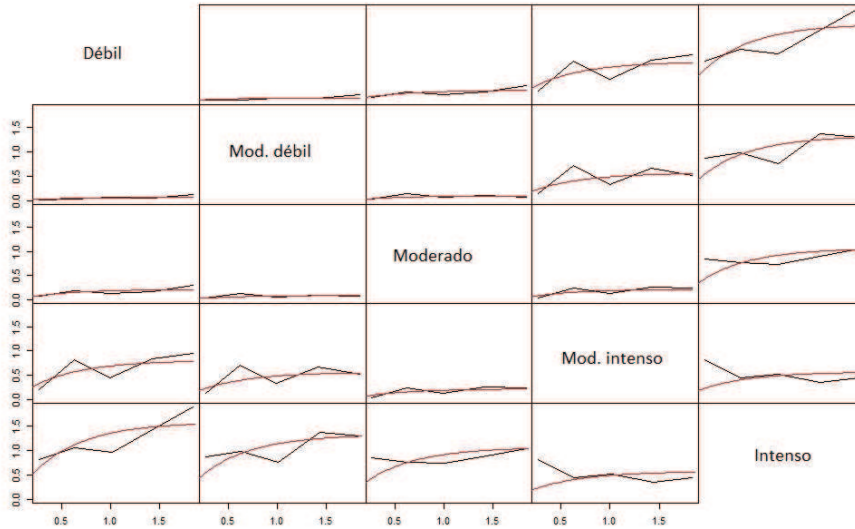


Figura 8.1: Variogramas de los log-ratios de cada par de variables (i, j) de los datos composicionales.

8.2. Análisis cluster

Para evaluar la similitud entre dos nodos SIMAR situados en dos posiciones diferentes, x y z , se utiliza la distancia Mahalanobis definida:

$$d(x, z) = (p_x - p_z)' C^{-1} (p_x - p_z). \quad (8.4)$$

donde p es el vector de probabilidades de cada categoría de temporales y C es la matriz de covarianzas:

$$C_{ij} = \frac{1}{N} (p_i - \mu_i)' (p_j - \mu_j). \quad (8.5)$$

Para tener en cuenta la distribución espacial de los nodos se sustituye la matriz de covarianzas por otra matriz positiva definida que tenga en cuenta la estructura espacial de los datos:

$$d(x, z) = (ilr(p_x) - ilr(p_z))' (2C(0) - \Gamma(h))^{-1} (ilr(p_x) - ilr(p_z)). \quad (8.6)$$

Donde ilr es la transformación 6.8 del simplex al espacio real, $C(0)$ es la función 8.3 definida en el apartado anterior y $\Gamma(h)$ tiene la siguiente expresión:

$$2\Gamma(h) = 2C(0) - (C(h) + C^t(h)). \quad (8.7)$$

De esta forma la distancia para muestras espacialmente dependientes será menor. Para h grande las variables son espacialmente independientes, la distancia converge a la distancia Mahalanobis clásica, mientras que para $h=0$ la distancia es la mitad que la clásica.

1. h grande: $\lim_{h \rightarrow \infty} \Gamma(h) = C(0)$, $\lim_{h \rightarrow \infty} C(h) = 0$

$$(2C(0) - \Gamma(h))^{-1} \rightarrow C(0)^{-1}$$

2. $h=0$: $\Gamma(0) = 0 \rightarrow (2C(0) - \Gamma(0))^{-1} = 1/2C(0)^{-1}$

Para definir los clúster se utiliza el método de Ward. Inicialmente cada objeto forma un grupo independiente y a continuación los objetos o grupos próximos se juntan sucesivamente. El dendrograma resultante se puede ver en el apartado siguiente y dependiendo el nivel de similitud que se escoja se forman más o menos grupos.

Capítulo 9

Resultados

En este capítulo se muestran los resultados obtenidos con el software R (ver R Development Core Team (2008)). El análisis de los datos composicionales se ha hecho utilizando la librería *compositions* (ver van den Boogaart *et al.* (2014)). Asimismo, para obtener las muestras de la distribución *posteriori* se ha utilizado RStan. El proyecto Stan desarrolla un lenguaje de programación que implementa inferencia estadística bayesiana a través del algoritmo de Hamilton Monte Carlo. Este método tiene ventajas respecto otros, como el de Gibbs por ejemplo, ya que converge a la distribución objetivo más rápidamente. La diferenciación automática se utiliza para evaluar rápidamente y con precisión los gradientes sin la necesidad de que el usuario calcule las derivadas parciales. RStan es una herramienta reciente que aporta calidad a las estimaciones por el algoritmo que utiliza (para más información ver Stan Development Team (2015)).

9.1. Regionalización de los datos originales

Se ha obtenido 4 grupos como se observa en el dendrograma 9.1.

Los nodos SIMAR, numerados, y las regiones en las que se ha dividido el área mediterránea se pueden ver identificadas en diferentes colores en la imagen 9.2.

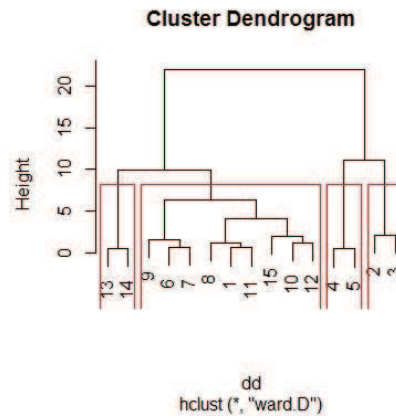


Figura 9.1: Dendrograma de los 4 clústers obtenidos de los datos composicionales iniciales.

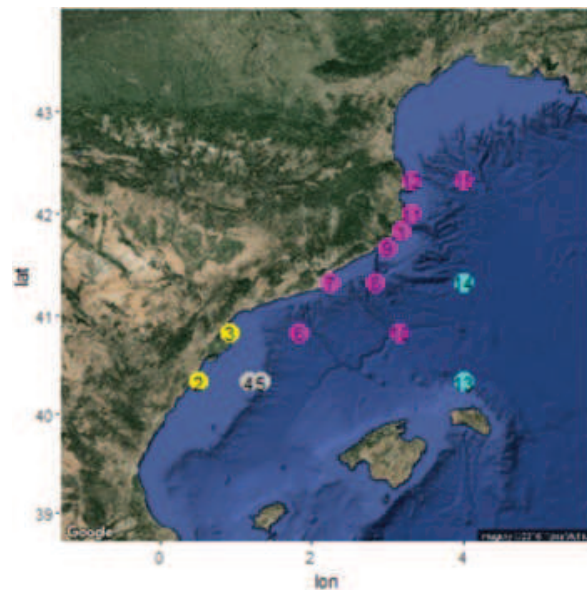


Figura 9.2: Regionalización en 4 clases a partir de los datos originales. (Fuente: elaboración propia basada en Google Maps)

Como se puede ver en el dendrograma, en primer lugar se separa los nodos del delta del Ebro (en amarillo en la figura 9.2) ya que son los que tienen menos similitudes con el resto. Esto puede ser debido a las diferencias climáticas y a

las circulaciones marítimas del delta del Ebro. Las aguas del Ebro, al entrar en contacto con la corriente general que se dirige hacia el SW siguiendo el talud continental, se ven frenadas y mezcladas con las aguas densas que han aflorado sobre la plataforma. En parte se ven atrapadas en un remolino anticiclónico en el golfo de Sant Jordi y mayoritariamente son desviadas hacia el sur. En el golfo de Valencia se ven desplazadas hacia la costa, al mismo tiempo que núcleos de esta agua poco densa se separan del flujo principal y generan movimientos de carácter anticiclónico en la zona del talud.

Dentro de los dos grupos formados, nodos del delta y el resto, se separan los nodos cercanos a la costa y los nodos alejados de la costa. En el grupo del sur se separan los nodos 2, 3 y 4, 5. En el otro grupo se separan los nodos 13 y 14 (en cian en la imagen 9.2) de los demás, situados en el litoral. En los nodos número 13 y 14 se observa que las alturas de ola son superiores a las de los demás nodos. Debido a este motivo estos dos nodos quedan agrupados formando el clúster representado en cian en la figura 9.2.

9.2. Regionalización de las muestras de las distribuciones posteriori

A partir de la información que contienen las muestras simuladas se ha obtenido 5 grupos, como se observa en el siguiente dendrograma 9.3.

Los nodos SIMAR, numerados, y las regiones en las que se ha dividido el área mediterránea se pueden ver identificadas en diferentes colores en la imagen 9.4.

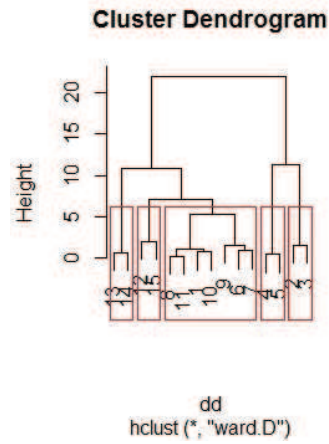


Figura 9.3: Dendrograma de los 5 clústers obtenidos a partir de las muestras del posteriori. Las líneas vericales representan en nivel de disimilaridad entre grupos.

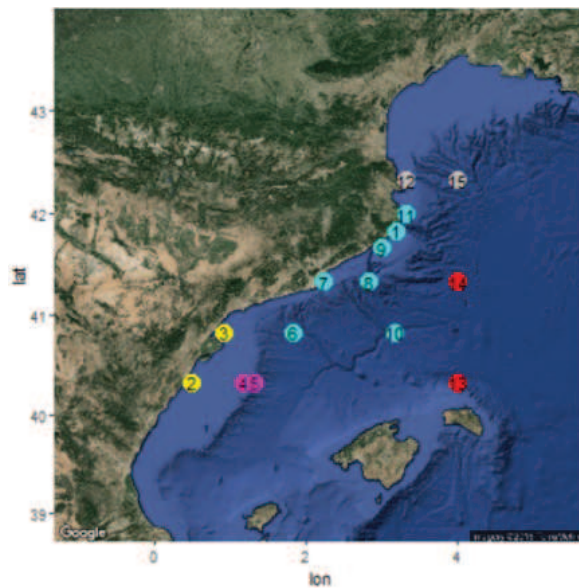


Figura 9.4: Regionalización en 5 clases a partir de la muestra simulada. (Fuente: elaboración propia basada en Google Maps)

En la regionalización a partir de las muestras simuladas se observa que los resultados son muy parecidos a los obtenidos a partir de los datos composicionales iniciales. Primero se separan los nodos del delta del Ebro (2, 3, 4, y 5 de la figura

9.4) del resto y después en cada grupo se separan los nodos cercanos a la costa de los de aguas profundas (nodos 13 y 14 en rojo en la figura 9.4). En este caso los nodos 12 y 15 (en gris en la figura), situados en el golfo de Roses se separan formando un nuevo grupo. En total se forman 5 clústers.

9.3. Muestras, en coordenadas, de las distribuciones posteriori

A continuación, se muestran los resultados obtenidos en cinco nodos, uno para cada clúster de la regionalización obtenida a partir de las muestras simuladas. Las expresiones de las coordenadas, interpretadas como proporciones entre categorías de temporales eran las siguientes:

$$c_1 = \ln \left[\frac{(x_1 x_2 x_3) \sqrt{2/15}}{(x_4 x_5) \sqrt{3/10}} \right]. \quad (9.1)$$

$$c_2 = \ln \left[\frac{(x_4) \sqrt{1/2}}{(x_5) \sqrt{1/2}} \right]. \quad (9.2)$$

$$c_3 = \ln \left[\frac{(x_1 x_2) \sqrt{1/6}}{(x_3) \sqrt{2/3}} \right]. \quad (9.3)$$

$$c_4 = \ln \left[\frac{(x_1) \sqrt{1/2}}{(x_2) \sqrt{1/2}} \right]. \quad (9.4)$$

En las figuras 9.9, 9.12, 9.5, 9.15 se observan los histogramas de la coordenada 1, que es el balance entre temporales débiles y moderados frente a los extremos. Se observa que después de 1989, este balance tiene valores más pequeños. En las figuras 9.10, 9.13, 9.6, 9.16 se observan los histogramas de la coordenada 2, que es el balance entre temporales moderadamente extremos frente a los extremos. En las figuras 9.11, 9.14, 9.7, 9.17 se observan los histogramas de la coordenada 3, que es el balance entre temporales débiles frente a los moderados. En la figura 9.8 se observan los histogramas de la coordenada 4, que es el balance entre temporales débiles frente a los moderadamente débiles. Se observa que después de 1989, este balance se mantiene más o menos estable en todos los nodos.

Para el nodo SIMAR 2132124, correspondiente al grupo rojo de la figura 9.4, los histogramas obtenidos son los de las figuras 9.5, 9.6, 9.7 y 9.8.

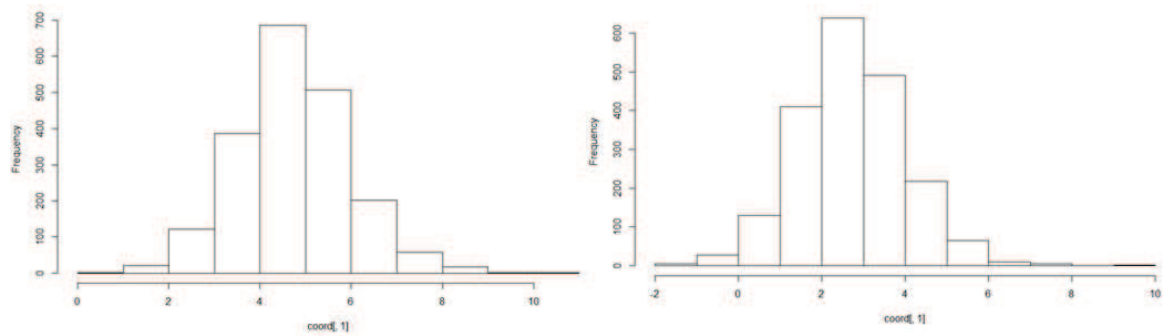


Figura 9.5: Histogramas de la coordenada 1 del nodo SIMAR 2132124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

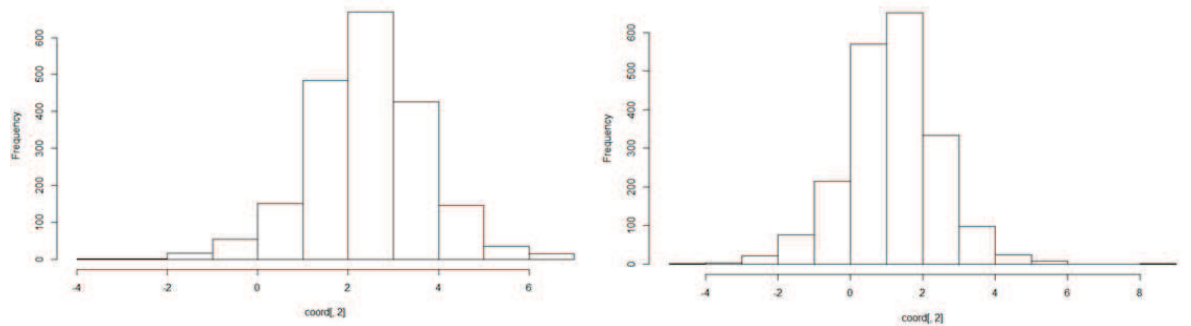


Figura 9.6: Histogramas de la coordenada 2 del nodo SIMAR 2132124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

CAPÍTULO 9. RESULTADOS

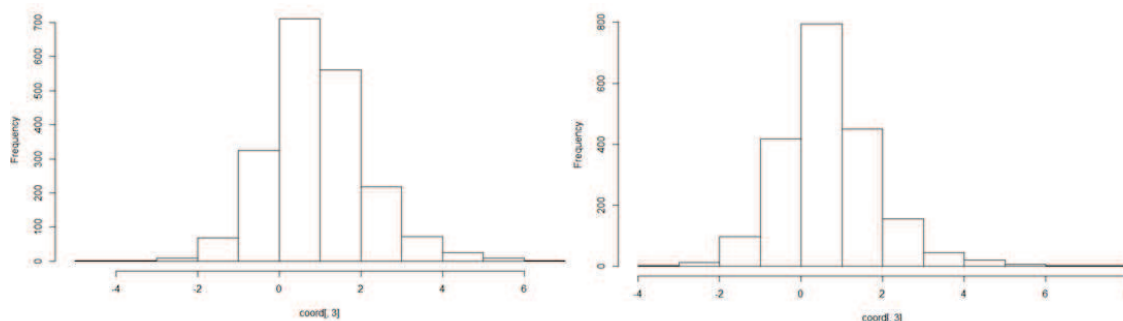


Figura 9.7: Histogramas de la coordenada 3 del nodo SIMAR 2132124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

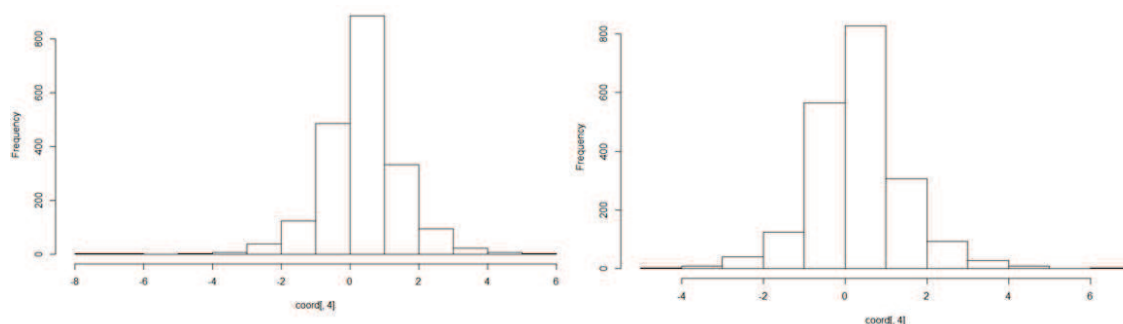


Figura 9.8: Histogramas de la coordenada 4 del nodo SIMAR 2132124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

El comportamiento de este nodo es el que se espera del cambio climático según Casas-Prat & Sierra (2013), las coordenadas 1 y 2 disminuyen (figuras 9.5 y 9.6) mientras que la 3 y 4 se mantienen estables (9.7 y 9.8). A continuación se compara los resultados obtenidos en el resto de clústers con los de este nodo.

Para el nodo SIMAR 2100124, correspondiente al grupo rosa de la figura 9.4, los resultados obtenidos son los siguientes.

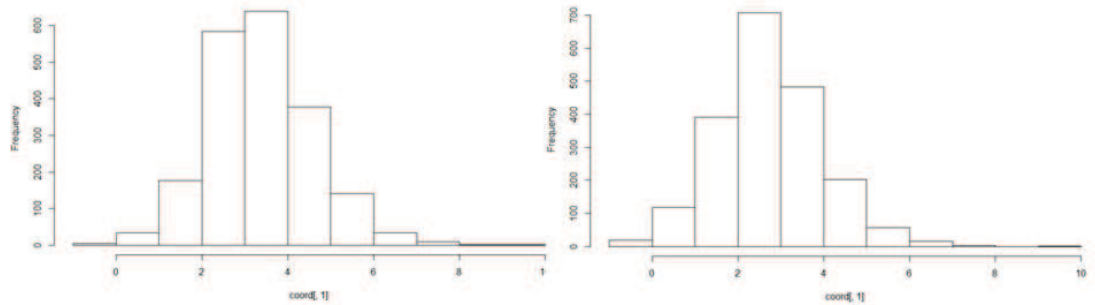


Figura 9.9: Histogramas de la coordenada 1 del nodo SIMAR 2100124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

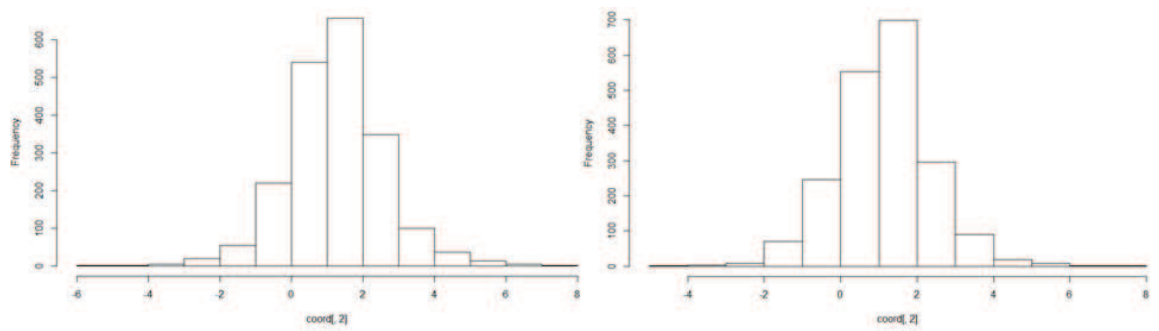


Figura 9.10: Histogramas de la coordenada 2 del nodo SIMAR 2100124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

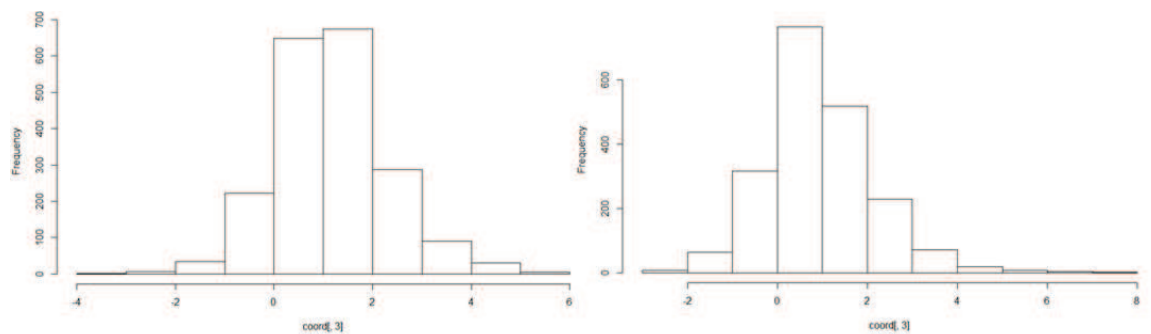


Figura 9.11: Histogramas de la coordenada 3 del nodo SIMAR 2100124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

CAPÍTULO 9. RESULTADOS

Como se observa en la figura 9.9 la primera coordenada disminuye en el segundo periodo de tiempo como en el nodo 2132124, mientras que la coordenada 2 se mantiene estable (9.10) y la tercera disminuye ligeramente (9.11), a diferencia de lo que se observaba en el nodo 2132124.

Para el nodo SIMAR 2122130, correspondiente al grupo cian de la figura 9.4, los resultados muestran un comportamiento similar al nodo 2100124.

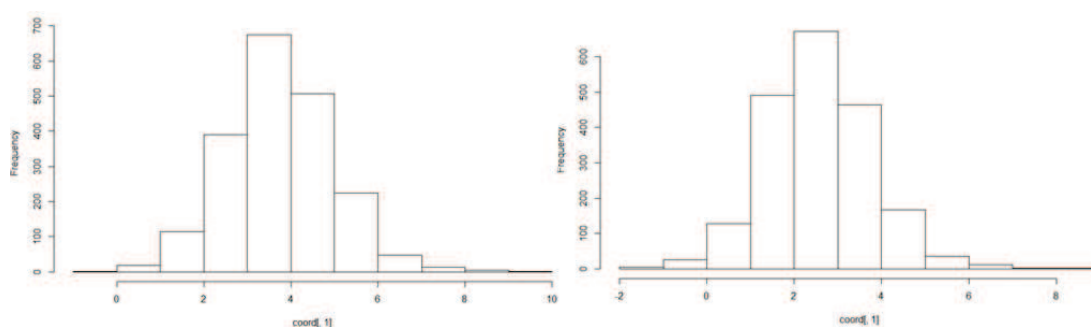


Figura 9.12: Histogramas de la coordenada 1 del nodo SIMAR 2122130. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

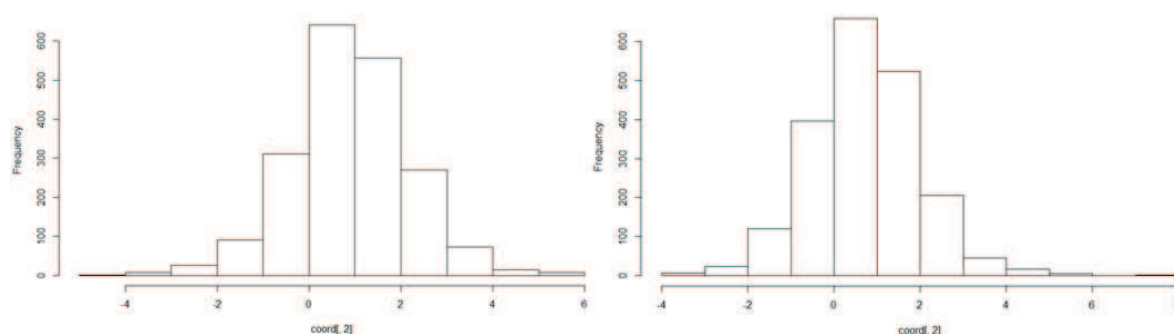


Figura 9.13: Histogramas de la coordenada 2 del nodo SIMAR 2122130. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

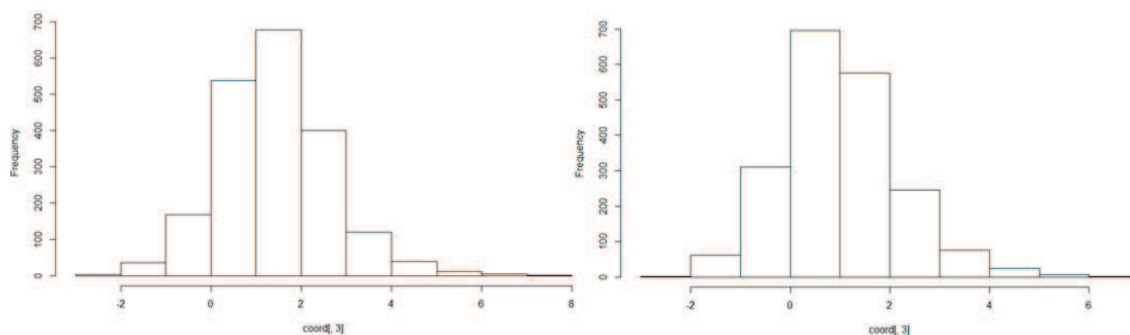


Figura 9.14: Histogramas de la coordenada 3 del nodo SIMAR 2122130. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

Para el nodo SIMAR 2132148, correspondiente al grupo gris de la figura 9.4, los resultados son similares a los dos casos anteriores.

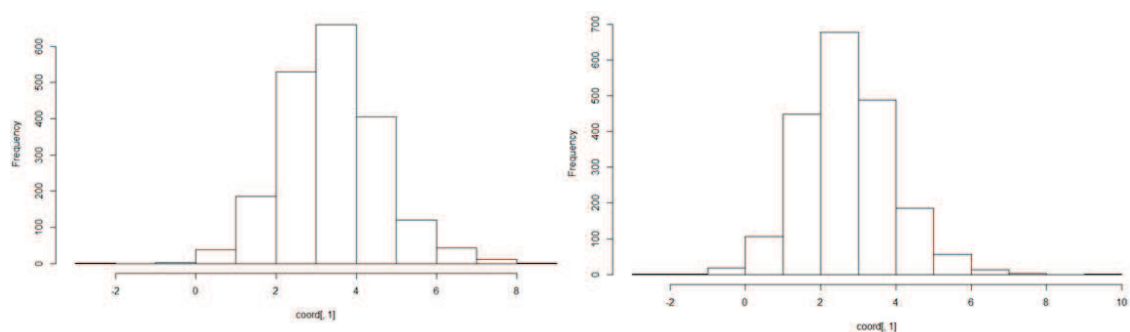


Figura 9.15: Histogramas de la coordenada 1 del nodo SIMAR 2132148. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

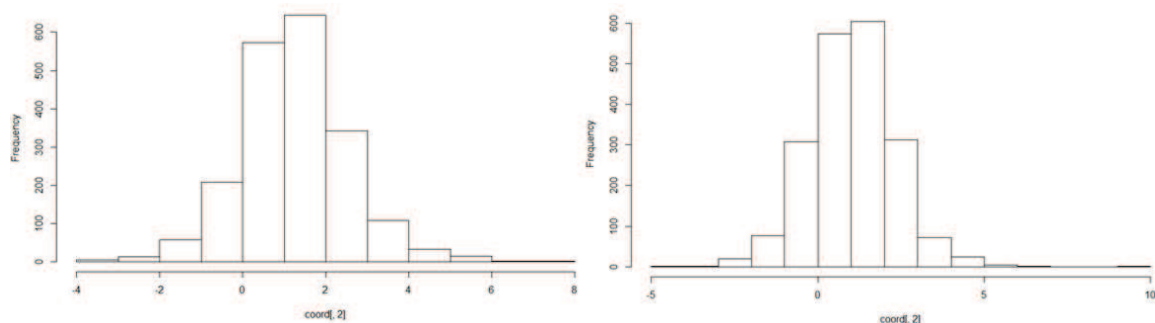


Figura 9.16: Histogramas de la coordenada 2 del nodo SIMAR 2132148. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

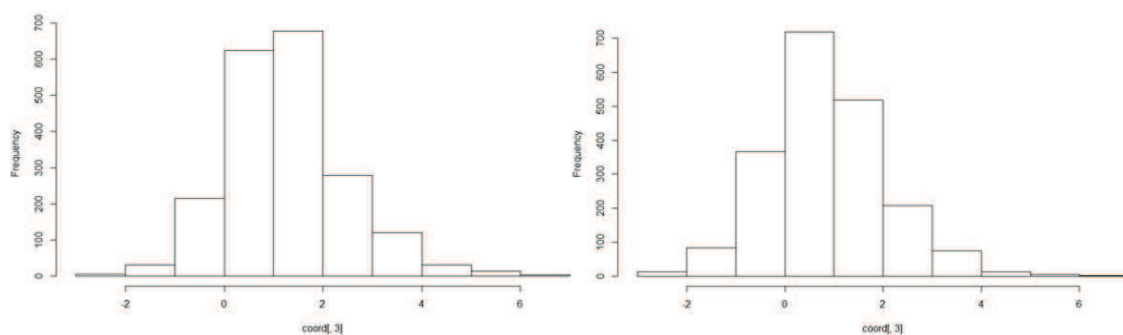


Figura 9.17: Histogramas de la coordenada 3 del nodo SIMAR 2132148. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

Para el nodo SIMAR 2090124, correspondiente al grupo amarillo de la figura 9.4, los resultados muestran que la coordenada 1 y 2 (figuras 9.18 y 9.19) aumentan a partir de 1988. Este comportamiento es diferente al resto de clústers, donde estos valores disminuían.

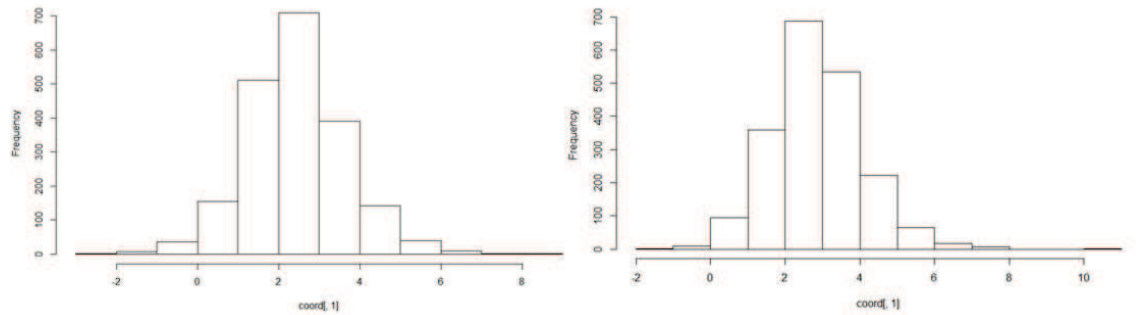


Figura 9.18: Histogramas de la coordenada 1 del nodo SIMAR 2090124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

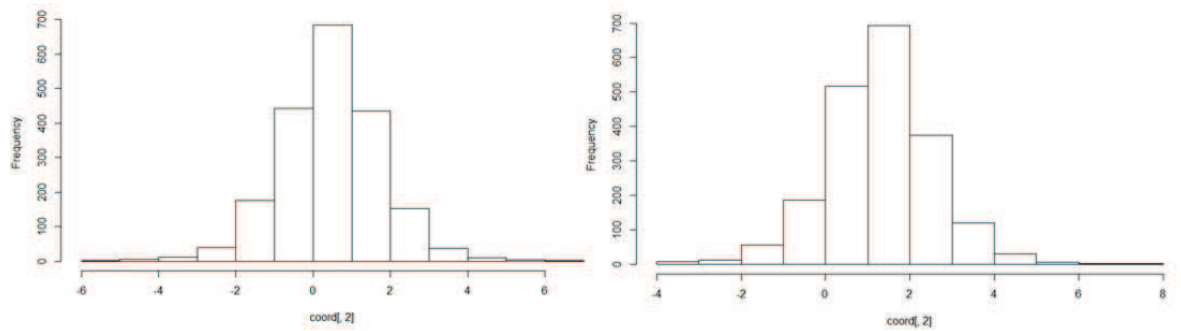


Figura 9.19: Histogramas de la coordenada 2 del nodo SIMAR 2090124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

En este capítulo se han presentado los resultados que aportan más información. El resto de histogramas se presentan en el Apéndice A.

Capítulo 10

Conclusiones

El potencial aumento en la intensidad de los temporales a causa del cambio climático es un problema fundamental en la ingeniería de costas ya que afecta al diseño y a la seguridad de obras marítimas. Por este motivo, el objetivo de este trabajo ha sido estudiar si resulta verosímil afirmar que ha habido un aumento de la probabilidad de los sucesos extremos.

En primer lugar, se ha caracterizado el clima extremal ajustando distribuciones de extremos como la distribución de Weibull y la distribución Generalizada de Pareto. A partir de los datos de temporales en 15 nodos SIMAR, entre 1958 y 2015, se ha establecido 5 categorías de temporales en función de su altura de ola y se ha estimado las probabilidades de cada categoría para ver si se mantienen estables o ha habido una variación. Para afrontar el problema se ha desarrollado un método de estimación de proporciones basado, por una parte, en la geometría de Aitchison en el simplex y en la correspondiente transformación *ilr* a coordenadas y, por otra, en técnicas de inferencia bayesianas mediante el uso de métodos de Montecarlo. Para la simulación de la muestra *a posteriori* con el método de Hamilton Monte Carlo se ha utilizado RStan, una herramienta innovadora que aporta calidad a las estimaciones por el algoritmo que utiliza. Con la información de las muestras simuladas de la distribución *a posteriori* se ha clasificado los 15 nodos SIMAR en clústers.

En cuanto a la agrupación en clústers de los nodos a partir de los datos composicionales iniciales se pueden distinguir 4 grupos. Como se puede ver en el dendrograma, en primer lugar se separa los nodos del delta del Ebro del resto.

Esto puede ser debido a las diferencias climáticas y a las circulaciones marítimas del delta del Ebro. Dentro de los dos grupos formados se separan los nodos cercanos a la costa y los nodos alejados de la costa. En la regionalización a partir de las muestras simuladas se observa que los nodos situados en el golfo de Roses se separan formando un nuevo grupo. En total se forman 5 clústers.

En los clústers representados en la figura 9.4 se ha estudiado las muestras de los nodos SIMAR 2100124,2122130,2132124 y 2132148. A partir del análisis estadístico realizado se concluye que el comportamiento de los clústers es, en general, el que se atribuye al cambio climático, los temporales extremos aumentan mientras que los medios se mantienen más o menos estables. El nodo en el que se observa más claramente este comportamiento es el 2132124. Se observa que la primera y segunda coordenada son las que más disminuyen en el periodo 1989-2015 respecto al 1958-1988, mientras que las coordenada 3 y 4, relativas al clima medio, se mantienen relativamente estables. La primera coordenada es el balance entre los temporales débiles y los intensos, como se recuerda en la ecuación 9.1. Una disminución en esta coordenada indica que los temporales intensos han aumentado frente a los moderados y débiles. La segunda coordenada, ecuación 9.2, representa la ratio entre los temporales moderadamente intensos y los intensos. La disminución de esta coordenada indica el aumento de los sucesos extremos.

El nodo SIMAR 2090124, perteneciente al clúster rosa de la figura 9.4, situado en el sur de la costa catalana, es el único que presenta un comportamiento diferente al resto. Se observa que el valor de la primera coordenada ha aumentado en los últimos 26 años, mientras que en los otros nodos disminuye. Este resultado es coherente con la agrupación representada en el dendrograma 9.3, donde se observa que los nodos del delta del Ebro son los que más difieren del resto.

Se tiene que tener en cuenta que este estudio tiene una limitación importante debida a que solo se dispone de datos a partir del año 1958 y para poder afirmar que ha habido un cambio en el clima marítimo sería necesario disponer de datos pertenecientes a un periodo de tiempo más extenso.

Capítulo 11

Trabajos futuros

Para la realización de este trabajo se ha desarrollado un método de estimación de proporciones basado, por una parte, en la geometría de Aitchison en el simplex y en la correspondiente transformación *ilr* a coordenadas y, por otra, en técnicas de inferencia bayesianas mediante el uso del método de Hamilton Montecarlo. Este método se podía utilizar para estimar el vector de probabilidades multinomiales en cualquier problema de datos en variables categóricas. A partir de los resultados de las muestras de las distribuciones *a posteriori* se podría desarrollar un estudio más exhaustivo de comparación de proporciones para confirmar si la hipótesis de cambio de clima resulta verosímil. Además, en el estudio del clima también se debería incluir otros parámetros que se tienen en cuenta en el diseño de obras marítimas, tales como la duración o periodo de pico de los temporales.

Capítulo 12

Agradecimientos

En primer lugar quiero mostrar mi agradecimiento a los tutores, Maribel Ortego y Jesús Corral, por su ayuda y orientación durante la realización de este trabajo que me ha permitido aumentar los conocimientos básicos de estadística adquiridos durante la carrera. Asimismo, agradezco a Puertos del Estado por cederme los datos que han hecho posible la realización de este estudio.

Bibliografía

- (2001). *ROM 0.0. Procedimiento general y bases de cálculo en el proyecto de obras marítimas y portuarias*. Ministerio de Fomento-Puertos del Estado.
- (2013). *Climate change adaptation, coastal and marine issues*. European Commission.
- (2015). *Stan Modeling Language. User's Guide and Reference Manual*. Stan Development Team.
- (2016). *www.puertos.es/es – es/oceanografia/Paginas/portus.aspx*. Puertos del Estado (b), consultada en octubre 2015.
- (2016). *www.puertos.es/oceanografia_y_meteorologia/redes_de_medida/index.html*. Puertos del Estado (a), consultada en enero 2016.
- ALBERT, J. (2007). *Bayesian Computation with R*. Springer.
- CARLIN, B. & LOUIS, T. (2009). *Bayesian methods for data analysis*. Chapman andHall/CRC Press, 3 ed.
- CASAS-PRAT, M. & SIERRA, J. (2013). *Projected future wave climate in the NW Mediterranean Sea*. Journal of Geophysical Research Oceans.
- EVERITT, B. S. (1993). *Cluster Analysis*. Cambridge (UK): Edward Arnold.
- GELMAN, A., CARLIN, J., STERN, H., DUNSON, D., VEHTARI, A. & RUBIN, D. (2014). *Bayesian Data Analysis*. Chapman andHall/CRC Press, 3 ed., pp. 276,589–606.
- GISBERT MARTÍN DE HIJAS, X. (2007). *Estimación Bayesiana de proporciones: Aplicación a la detección de cambios de clima*. UPC, ETSECCPB.

- HARLEY, C. D. G., HUGHES, A. R., HULTGREN, K. M., MINER, B. G., SORTE, C. J. B., THORNBUR, C. S., RODRIGUEZ, L. F., TOMANEK, L. & WILLIAMS, S. L. (2006). The impacts of climate change in coastal marine systems. *Ecology Letters* **9**, 228–241.
- IPCC (2007). Cuarto informe del grupo intergubernamental de expertos sobre el cambio climático.
- IPCC (2012). Managing the risks of extreme events and disasters of advance climate change adaptation.
- KRUSCHKE, J. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Elsevier.
- ORTEGO, M., GIBERGANS-BÁGUENA, J. & TOLOSANA-DELGADO, R. (2011). *Pluviometric regionalization of Catalunya: a compositional data methodology*. Proceedings of the 4th International Workshop on Compositional Data Analysis, St. Feliu de Guíxols, Girona (Spain).
- PAWLOWSKY-GLAHN, V., EGOZCUE, J. J. & TOLOSANA-DELGADO, R. (2015). *Modeling and Analysis of Compositional Data*. Wiley, 1 ed., pp. 32–58,130–146.
- PAWLOWSKY-GLAHN, V. & OLEA, R. (2004). *Geostatistical analysis of compositional data*. Oxford University Press.
- PICKANDS, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119–131.
- R DEVELOPMENT CORE TEAM (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- ROBERT, C. & CASELLA, G. (2004). *Monte Carlo Statistical Methods*. New York, USA: Springer Verlag, 2 ed.
- SÁNCHEZ ARCILLA, A. & GONZÁLEZ-MARCO (2008). A review of wave climate and predictions along the spanish mediterranean coast. *Nature Hazards and Earth Systems Science* **8**(6), 1217–1228.
- TOLOSANA, R., VAN DEN BOOGAART, K. G. & PAWLOWSKY-GLAHN, V. (2011). Geostatistics for compositions. In: *Compositional Data Analysis at the beginning of the XXI century* (PAWLOWSKY-GLAHN, V. & BUCCIANTI, A., eds.). Wiley.

BIBLIOGRAFÍA

VAN DEN BOOGAART, K. G., TOLOSANA, R. & BREN, M. (2014). *Compositional Data Analysis*. URL <http://www.stat.boogaart.de/compositions>. R package version 1.10-2.

VAN DEN BOOGAART, K. G. & TOLOSANA-DELGADO, R. (2013). *Analyzing Compositional Data with R*. Springer.

Apéndice A

Histogramas de las muestras posteriores

En este apéndice se presentan los histogramas de las muestras de la cuarta coordenada de los nodos SIMAR 2100124, 2122130 y 2132148, correspondientes a los grupos rosa, cian y gris de la figura 9.4. También aparecen los histogramas de la tercera y cuarta coordenada (figura A.4 y A.5) para el nodo SIMAR 2090124, correspondiente al grupo amarillo de la figura 9.4.

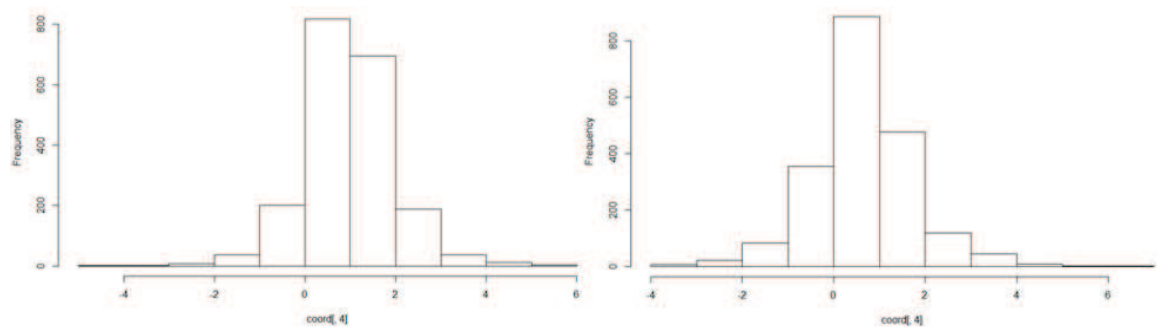


Figura A.1: Histogramas de la coordenada 4 del nodo SIMAR 2100124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

APÉNDICE A. HISTOGRAMAS DE LAS MUESTRAS POSTERIORI

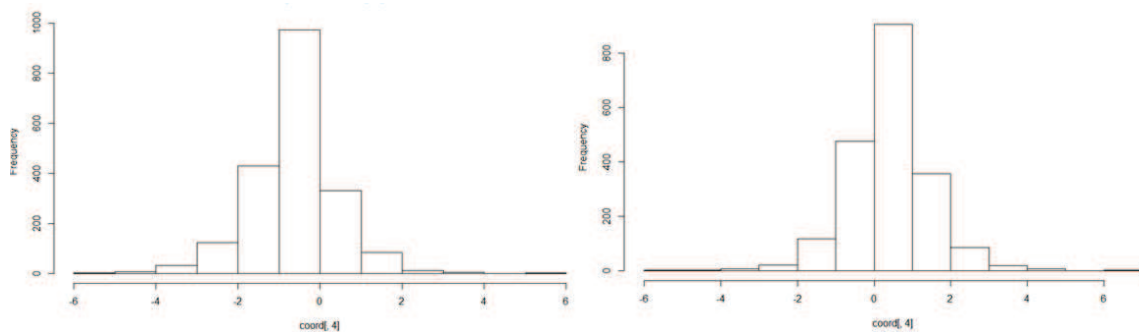


Figura A.2: Histogramas de la coordenada 4 del nodo SIMAR 2122130. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

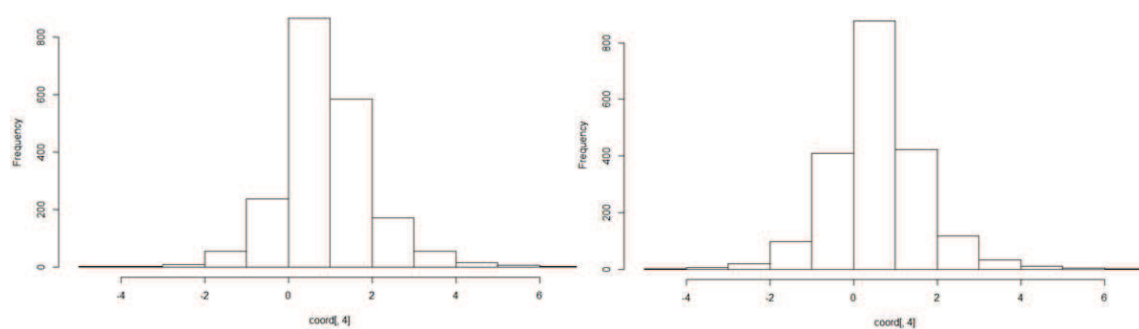


Figura A.3: Histogramas de la coordenada 4 del nodo SIMAR 2132148. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

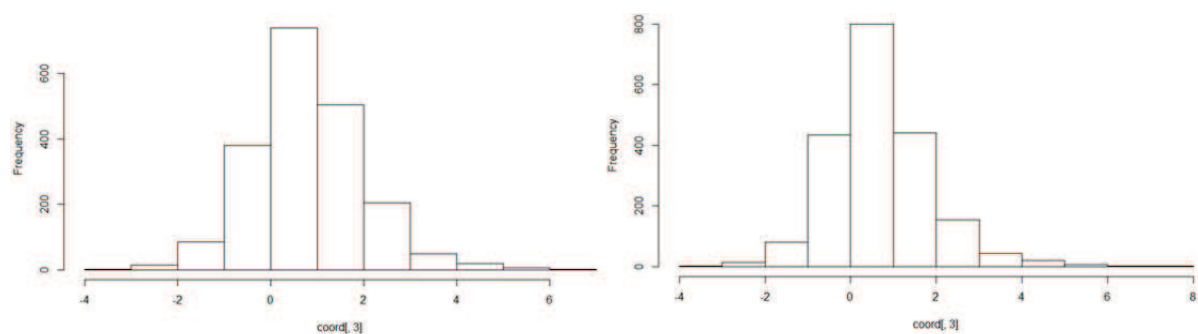


Figura A.4: Histogramas de la coordenada 3 del nodo SIMAR 2090124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.

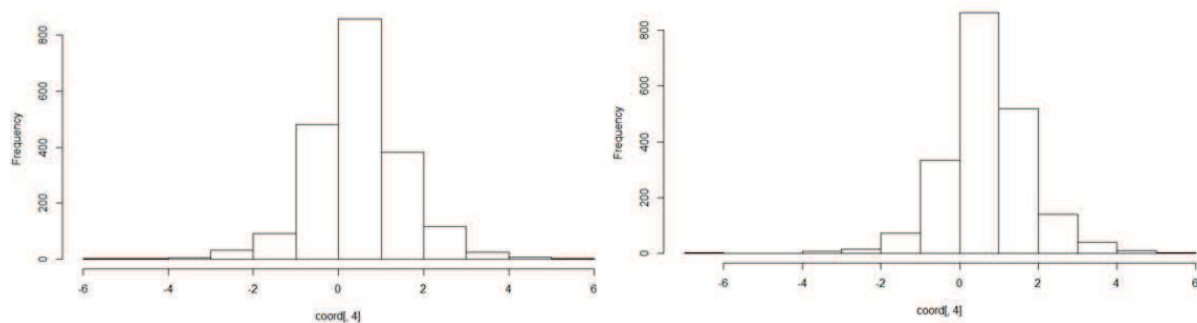


Figura A.5: Histogramas de la coordenada 4 del nodo SIMAR 2090124. Periodo 1958-1988 a la izquierda y 1989-2015 a la derecha.